Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Concept Attainment: I. The Role of Age and Intelligence in Concept Attainment by Induction: S. F. OSLER AND M. W. FIVEL 1

Concept Attainment: II. Effect of Stimulus Complexity upon Concept Attainment at Two Levels of Intelligence: S. F. OSLER AND G. E. TRAUTMAN 9

Familiarization (*n*) as a Stimulus Factor in Paired-Associate Verbal Learning: D. R. GANNON AND C. E. NOBLE 14

Paired-Associates Learning with Varying Relative Percentages of Occurrence of Alternative Response Members: A. E. GOSS AND M. E. SUGERMAN 24

Separating Perceptual and Linguistic Effects of Context Shifts upon Absolute Judgments: D. L. KRANTZ AND D. T. CAMPBELL 35

Effects of Foreperiod, Foreperiod Variability, and Probability of Stimulus Occurrence on Simple Reaction Time: D. H. DRAZIN 43

Free Recall and Ordering of Trigrams: L. M. HOROWITZ 51

Forward, Backward, and Pseudoconditioning of the GSR: R. A. CHAMPION AND J. E. JONES 58

Acquisition of Perceptual Responses as a Function of Loading, Location, and Repetition: H. W. KARN AND L. W. GREGG 62

Associative Factors in Verbal Transfer: J. BASTIAN 70

The Influence of Thirst and Schedules of Reinforcement-Nonreinforcement Ratios upon Brightness Discrimination: R. LACHMAN 80

Generalization Gradients in a Discrimination Situation: D. LABERGE 88

Supplementary Report: Effect of Practice on an Illusion: D. RETHLINGSHAFER AND T. I. SHERRER 95

American Psychological Association

Vol. 62, No. 1 July 1961

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1961 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1961 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 62, NO. 1 JULY 1961

CONCEPT ATTAINMENT: I. THE ROLE OF AGE AND INTELLIGENCE IN CONCEPT ATTAINMENT BY INDUCTION[1]

SONIA F. OSLER AND MYRNA WEISS FIVEL[2]

Goucher College

Department of Pediatrics, Johns Hopkins University

Most concept studies with children have utilized either the interview-questionnaire method or the performance method (Vinacke, 1952, p. 108). Each of these procedures presents certain obstacles to the discovery of *S*'s knowledge of a specific concept or his ability to utilize it. Implicit in the interview-questionnaire method, illustrated in the early works of Piaget (1929, 1930) and others (Klingberg, 1957; Russel, 1940, 1942), is the assumption that knowledge of a concept is necessarily accompanied by the child's ability to translate it into words. That this assumption is not justified has been demonstrated in a variety of ways with adults and children (Hazlitt, 1930; Heidbreder, 1927; Hull, 1920). It is primarily for this reason that Piaget modified his early procedures involving only interviews to the combined interview performance technique.

With few exceptions the performance procedures utilized in concept studies boil down essentially to a type of sorting technique. The tasks may vary from matching to stimulus (Brian & Goodenough, 1929; Colby & Robertson, 1942) to sorting into categories that "go together" (Thompson, 1941; Welch, 1940). The difficulty associated with this technique is that performance may be a function

[1] This investigation was supported by a grant from the National Institute of Mental Health, Psychopharmacology Service Center, made to Leon Eisenberg of Johns Hopkins University, and the senior author (Grant MY-2583).

The writers wish to express their gratitude to Mary A. Adams, Assistant Superintendent, Elementary Education, Lorne H. Woollatt, Former Director of Bureau of Research, Vernon S. Vavrina, Director, Junior and Senior High Schools, and to Arthur Lichtenstein, Director of Special Services for Pupils, Baltimore City Department of Education; and to B. Melvin Cole, Director, Elementary Education, Norris A. King, Director, Secondary Education, and Charles J. Leiman, Supervisor, Psychological Services, Baltimore County Department of Education; and to the the principals and their staffs who cooperated in the conduct of the study.

They also wish to thank Grace E. Trautman, Nancy A. Freydberg, and Sandra J. Weiss for testing the subjects and for many suggestions concerning the conduct of the study; and to Natalia Chapanis for assistance with the statistical analysis of the data. To Jerome Cornfield and Margaret P. Martin go our special thanks for their invaluable assistance with the statistical problems encountered in this investigation.

[2] Now at the University of Pennsylvania.

of factors other than knowledge of the specific concept under investigation. For example, Heidbreder (1947) has demonstrated with adult *S*s a dominance hierarchy in concept learning and sorting, in which object, form, and number concepts occupy in this order decreasing positions in the hierarchy; and Brian & Goodenough (1929) have similarly demonstrated with children the relative dominance of color and form. Others have shown that specific stimulus attributes may affect their relative dominance (Dattman & Israel, 1951; Heidbreder, 1952; Stevenson & McBee, 1958). Furthermore, it is reasonable to assume that where alternative possibilities for categorizing exist the preference, interest, or experience of *S* will affect his performance.

It would, therefore, seem desirable to use a technique for concept studies which would avoid the difficulties associated with the interview-questionnaire and sorting techniques and yield a sensitive measure of *S*'s capacity to deal with the task. The inductive method seems to fulfill these requirements. In this method *S* is required to give the same response to differing stimuli which belong to a common category. The *S* demonstrates his attainment of the concept when he can respond appropriately to new stimuli at a level of accuracy specified by *E*. The stimuli are so designed that in order to achieve the required level of success *S* must categorize according to the concept expected by *E*. In this manner competition from other concepts is diminished, if not altogether eliminated. Furthermore, a correct response does not involve verbalization; but at the same time ability to verbalize may be studied as an additional dependent variable.

The inductive method has been used in concept formation studies with children (Heidbreder, 1928; Levy & Cuddy, 1956; Long, 1940; Welch & Long, 1942, 1943) and adults (Archer, Bourne, & Brown, 1955), but in most of these studies *S* was restricted to a relatively small number of trials. Welch and Long (1943), for example, allowed only 30 trials for a problem involving Mills' joint method, and then concluded that no *S* was able to solve problems of this type. In the study by Levy and Cuddy, a rote learning component was introduced in a modification of the inductive method in which only 10 different stimuli were presented over and over again. In the investigation described below 150 trials were given (unless success was achieved earlier). A pilot study revealed that only one or two *S*s out of 100 who do not succeed within 150 trials attain success if given additional opportunity. A sufficiently large number of trials thus provides the possibility not only for the study of success in attaining the concept but also for estimating the relative difficulty of the concept as a function of other variables, such as age and intelligence.

In fact, one of the aims of the present investigation was to determine the role of intelligence in concept attainment by induction. While it seems reasonable to assume that intelligence and concept attainment are highly correlated, such a relationship has not been consistently demonstrated (Eisman, 1958; Griffith, Spitz, & Lippman, 1959; Hoffman, 1955; Plenderleith, 1956; Stuart, 1935). Furthermore, most studies have compared normal and feebleminded *S*s, while in the present investigation a comparison is made between normal and superior *S*s. Three age groups were selected for this study: 6-, 10-, and 14-yr. *S*s. These age groups were

chosen because they represent the extremes and midpoint of the elementary school population and because they fall within three stages of mental development as defined by Piaget.

The concepts studied were: bird, animal, and living thing. According to the notions of Welch and Long (1940) these three concepts should represent increasing levels of difficulty, and thereby enhance the discriminating power of the tasks.

METHOD

Subjects.—The *S*s consisted of 180 school children attending elementary and junior high schools of Baltimore City and County, equally divided among the following age groups 6-, 10-, and 14-yr. Each *S* was within 6 mo. of the age specified. Half of the children within each age level had a WISC IQ between 90 and 109 (mean IQ for total normal intelligence group was 101.6, $SD = 5.1$) and the other half had a WISC IQ of 110 and above (mean IQ for total high intelligence group was 121.3, $SD = 8.1$).

Apparatus.—The apparatus consisted of two 3-in. hollow wooden cubes painted black. One side of the cube was missing so that it could be placed over a marble. A metal frame was attached to the side of the cube facing *S* in such a way that 2.75-in. cardboard stimuli could easily be slipped into it. A set of stimuli for each concept consisted of 150 pairs of pictures individually mounted on squares of white cardboard. One picture of each pair represented an example of the concept; the other picture was similar in area, color combination, and brightness, but represented an object not belonging to the concept category. A different set of stimulus pairs was used to represent each of the three concepts, bird, animal, and living thing, making a total of 900 stimulus cards. Figure 1 shows some of the stimulus pairs. It should be remembered, however, that the actual stimuli were multicolored.

The reason for choosing naturalistic stimuli rather than formal types which would allow for systematic variation of dimensions was to avoid the complication of unequal familiarity or availability of verbal labels, which would tend to favor the older and more intelligent *S*s. Prior to the actual investigation the stimulus material was shown to 10

FIG. 1. Samples of stimulus pairs used in the bird, animal, and living thing concepts, respectively.

children in each age and intelligence group to be tested. These *S*s were asked to identify the pictures on the cards and any picture which was not properly identified was discarded. The children used in the pretest did not serve as *S*s in the experiment. The familiarity of the material was further checked by asking each *S* to identify 10 pairs of stimuli upon completion of the concept learning task. Different sets of 10 were used for each *S*. These data confirmed the fact that the pictures were entirely familiar to all *S*s.

A black curtain, which was drawn closed between responses, was attached to a rod mounted on a frame and placed on the table between *E* and *S*. The *S* was provided with a grooved wooden tray for collecting the marbles with which his correct responses were reinforced.

Procedure.—Each *S* was tested individually in two separate sessions. At the first session *S* was told that he would play a game in which he could win any one of a group of toys displayed before him. He was asked to select the most preferred of the group, and this was then set slightly apart from the others in a position which allowed him to view it during

TABLE 1

MEAN NUMBER OF ERRORS PER GROUP FOR REACHING CRITERION OF 10 CONSECUTIVE CORRECT RESPONSES

Concepts	6-Yr. *Ss*				10-Yr. *Ss*				14-Yr. *Ss*				All Ages	
	Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ	High IQ
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*		
Bird	55.7	20.6	54.9	23.7	32.0	26.4	19.5	26.4	27.7	20.8	10.0	13.5	38.5	28.1
Animal	70.3	12.6	55.7	27.8	34.2	21.0	17.8	26.1	29.1	21.2	12.6	10.8	44.5	28.7
Living thing	66.2	5.9	50.9	23.0	47.6	21.2	38.7	22.0	30.8	28.2	9.4	10.5	48.2	33.0
All concepts	64.1		53.8		37.9		25.3		29.2		10.7		43.7	29.9

Note.—Each mean is based on 10 *Ss*, total $N = 180$.

the "game." The following instructions were then given:

> I am going to hide a marble under one of these boxes and your job is to find the marble. Sometimes the marble will be under this box, and sometimes it will be under this box (demonstrated). If you guess right, you may put the marble in your tray. If you don't guess right, I will keep the marble. I will put cards in here like this (demonstrated with sample cards unrelated to the concept studied). If you look at the cards, you will be able to get many marbles. When you get enough marbles to fill your tray you will get the (name of the toy).

The *S* was required to point to the box which he thought covered a marble. The *E* then lifted the box, and if the response was correct, *S* picked up the marble and placed it in his tray. The procedure was continued until *S* made 10 consecutive correct responses, at which point *E* commented on *S*'s good performance and asked him how he knew the location of the marble.[3] The maximum number of trials was 150; and any *S* who did not reach the criterion of success at that point was considered to have failed. Each *S* worked on one concept only. The second session, held approximately two weeks later, was devoted to the administration of the WISC.

These experiments were conducted in schools; and one serious problem that arose was communication among the children. To prevent this the number of *Ss* used in any one school was small; and *Ss*, who had been chosen at random from the appropriate age groups, were tested in sequences which were likely to minimize communication. A split-half comparison of scores was made and no significant difference resulted.

Experimental design.—A $3 \times 3 \times 2$ factorial design was used with the three age levels constituting one variable, the three concepts the second variable, and the two intelligence levels the third variable. This resulted in 18 groups with 10 *Ss* in each group, consisting of approximately the same number of boys and girls.

TABLE 2

ANALYSIS OF VARIANCE OF ERRORS

Source	*df*	*MS*	*F*
Concepts (C)	2	801.58	1.620
Age (A)	2	24,053.90	48.616***
Intelligence (I)	1	8,556.00	17.293***
C × A	4	753.76	1.523
C × I	2	135.84	—
A × I	2	274.24	—
C × A × I	4	150.73	—
Within	162	494.77	
Total	179		

*** $P < .001$.

RESULTS

The effect of age and intelligence.—The two response measures of errors to criterion and successful attainment

[3] Originally it was planned to ask each *S* periodically during the experimental session for the basis of his choice. However, we discovered that frequently *S*, after stating his reason (such as "things with two legs" when working with birds), would change his strategy anticipating *E* not to bait stimuli that he expected *S* to choose.

of criterion were analyzed. Table 1 shows the mean number of errors per group and the corresponding *SD*s. These data did not satisfy the requirement of homogeneity for an analysis of variance. Nevertheless, an analysis of variance was performed for exploratory purposes; and the results are shown in Table 2. Even with the caution dictated by the lack of homogeneity, the *P* values associated with the effects of age and intelligence gave strong indication that these variables produced significant differences in performance.

The nonparametric Kruskal-Wallis one-way analysis of variance by ranks (Siegel, 1956) was also performed, and these results, shown in Table 3, were in complete agreement with the analysis of variance. To check out a possible interaction between age and intelligence an analysis was done on the data collapsed across concepts, thus producing homogeneity of variance. The *F* ratio for the Age × Intelligence interaction was less than one.

The distribution of *S*s in the 18 experimental groups who achieved criterion performance is shown in Table 4. An analysis of these frequency data, shown in Table 5, yielded results completely consistent with the error scores.

As an additional measure of association, the rank-order correlation was computed between error scores and mental age scores. As shown in Table 6, the rho values obtained are the same for each concept group. The data were, therefore, pooled and the value of the correlation for the total *S* population computed. All the values obtained were significant at the .001 level.

TABLE 3

KRUSKAL-WALLIS ANALYSIS OF VARIANCE OF ERROR SCORES

Source	*H*	*df*	*P*
Age	58.86	2	.001
Intelligence	14.83	1	.001
Concepts	4.27	2	*ns*

The effect of concept hierarchy.—From the results it appears that concept hierarchy had no effect on one measure of performance, errors to criterion, and only a minimal effect on frequency of *S*s who achieved criterion performance. As shown in Table 5, the concepts did produce a significant difference in the Normal Intelligence group. This analysis was done according to Cochran (1954) to take advantage of the predictions of the differences to be expected. The *S*s of high intelligence found all concepts of equal difficulty.

DISCUSSION

That age and intelligence are associated with more effective concept learning

TABLE 4

NUMBER OF *S*S PER GROUP WHO REACHED CRITERION PERFORMANCE

Concepts	6-Yr. *S*s		10-Yr. *S*s		14-Yr. *S*s		Total	
	Normal IQ	High IQ	Normal IQ	High IQ	Normal IQ	High IQ	Normal IQ	High IQ
Bird	5	4	8	8	8	10	21	22
Animal	2	5	7	8	8	10	17	23
Living thing	0	4	5	8	7	10	12	22
Total	7	13	20	24	23	30	50	67

TABLE 5

CHI SQUARE ANALYSIS OF FREQUENCIES OF SUCCESSFUL *Ss*

Comparison	*df*	Chi Square
Age	2	42.636***
Intelligence	1	7.058**
Concept	1[a]	2.967
Concept (Normal IQ)	1[a]	5.468*
Concept (High IQ)	2	.116

[a] This calculation is based on Cochran's method, which involves loss of 1 *df* for the regression equation.
* $P < .02$.
** $P < .01$.
*** $P < .001$.

TABLE 6

RANK-ORDER CORRELATIONS BETWEEN MA AND ERROR SCORES

Group	rho	*t*	*df*
Bird	−.58	5.41***	58
Animal	−.60	5.71***	58
Living thing	−.63	6.14***	58
Total	−.61	10.16***	178

*** $P < .001$.

seems entirely reasonable, but the specific mechanism by which the effectiveness is enhanced has yet to be clarified. Is the greater effectiveness of the older and more intelligent *Ss* due merely to enhanced speed of learning, that is, to a greater increment in response tendency per trial, or is there also a qualitative difference in the learning process? To answer this question learning curves of individual *Ss* were examined for the purpose of determining whether concept attainment was gradual or sudden. A gradual acquisition of the concept would be consistent with the continuity theory of learning and sudden acquisition with the theory of learning by hypothesis testing. In order to classify *Ss* into gradual and sudden learners their performance just prior to the 10 consecutive correct responses was examined. Each *S* was assigned a score which represented the percentage of correct responses in the 10 trials preceding the final 10 (which were all correct). All *Ss* whose score fell below the median, which was 60%, were classified as sudden learners, and those whose scores fell above the median were classified as gradual learners.[4] Table 7 shows the distribution of gradual and sudden learners, together with a chi square analysis, which shows that the incidence of sudden learners was a function of intelligence, but that neither age nor concept hierarchy produced any dif-

[4] Five *Ss* who fell into the gradual category had each achieved the concept in 20 trials or less and made no more than two errors. These *Ss* were arbitrarily placed into the sudden category, as their performance seemed to justify. This accounts for the unequal number of *Ss* in the sudden gradual categories. Two of the *Ss* had normal and three had high intelligence.

TABLE 7

BREAKDOWN OF SUCCESSFUL *Ss* INTO GRADUAL AND SUDDEN LEARNERS

Concepts	6-Yr. *Ss*				10-Yr. *Ss*				14-Yr. *Ss*				All Ages			
	Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ		High IQ	
	S	G	S	G	S	G	S	G	S	G	S	G	S	G	S	G
Bird	1	4	3	1	3	5	6	2	2	6	7	3	6	15	16	6
Animal	2	0	2	3	4	3	7	1	4	4	7	3	10	7	16	7
Living thing	0	0	1	3	2	3	2	6	4	3	7	3	6	6	10	12
Total	3	4	6	7	9	11	15	9	10	13	21	9	22	28	42	25

Note.—S refers to sudden learning, G to gradual learning. Age $\chi^2 = 1.066$, $df = 2$, not significant; Intelligence $\chi^2 = 4.034$, $df = 1$, $P < .05$; Concepts $\chi^2 = 2.731$, $df = 2$, not significant.

ference. From this finding an association between intelligence and concept attainment by hypothesis testing was inferred; an inference which is entirely consistent with recent findings by Kendler and Kendler (1959) and Kendler, Kendler, and Wells (1960).

One other factor that may be related to performance must be considered. The instructions given to *Ss* did not direct them explicitly to try to discover a concept; but merely stated that attending to the stimuli would help them win marbles. Theoretically, therefore, it is possible that the experimental task consisted of two stages: (*a*) the discovery of the problem, that is, that a concept is to be sought, and (*b*) the solution of the problem, that is, the discovery of the specific concept. The differences in performance associated with age and intelligence may be due to more effective performance at either or both of these stages. To determine whether problem finding is a significant variable and the extent of its influence upon performance, if any, would require a systematic study of the effect of instructions upon concept attainment. Recent findings on the effects of stimulus complexity upon concept attainment at two levels of intelligence (Osler & Trautman, 1961) support the interpretation of a difference in the process of concept attainment as a function of intelligence.

A few brief remarks concerning the correlation between errors and mental age are in order. The rank-order correlation was selected because the data did not meet the assumptions for either the product moment or the correlation ratio. The actual value of rho obtained is probably lower than the true value because of the restricted range of talent in the present sample. The range of mental ages was restricted in terms of the ages studied and the minimum IQ required; the error scores were restricted because of the maximum number of trials permitted. The relatively high values of rho obtained, despite these limitations, emphasize further the importance of the variables studied.

Summary

This investigation was concerned with the effect of age and intelligence on concept attainment. The *Ss* were 180 children, 60 in each of the following age groups: 6, 10, and 14 yr. They were divided equally into normal and high intelligence groups. Each *S* worked on one of the following concepts: bird, animal, or living thing.

The response measures of trials to criterion and the number of successful *Ss* were analyzed. The following conclusions were drawn: (*a*) Age was associated with significant differences in errors to criterion and number of successful *Ss*. (*b*) Intelligence was associated with significant differences in errors to criterion and number of successful *Ss*. (*c*) The effect of concept hierarchy was nominal. It had no effect upon errors to criterion; but it did produce significant differences in number of *Ss* of normal intelligence who achieved the criterion of success. (*d*) When *Ss* were divided into sudden and gradual learners, it was found that the frequency of sudden learners was a function of intelligence, but not of age. This was interpreted to indicate the greater utilization of mediators by the superior *Ss*. (*e*) Rank-order correlations were computed between mental age and error scores and these were found to be highly significant.

References

Archer, E. J., Bourne, L. E., Jr., & Brown, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49**, 153–164.

Brian, C. R., & Goodenough, F. L. Relative potency of color and form perception at various ages. *J. exp. Psychol.*, 1929, **12**, 197–213.

Cochran, W. G. Some methods for strengthening the common χ^2 tests. *Biometrics*, 1954, **10**, 417–451.

Colby, M. G., & Robertson, J. B. Genetic studies in abstraction. *J. comp. Psychol.*, 1942, **33**, 385–401.

Dattman, P. E., & Israel, H. E. The order of dominance among conceptual capacities: An experimental test of Heidbreder's hypothesis. *J. Psychol.*, 1951, **31**, 147–160.

Eisman, B. S. Paired associate learning, generalization and retention as a function of intelligence. *Amer. J. ment. Defic.*, 1958, **63**, 481–489.

Griffith, B. C., Spitz, H. H., & Lipman, R. S. Verbal mediation and concept formation in retarded and normal subjects. *J. exp. Psychol.*, 1959, **58**, 247–255.

HAZLITT, V. Children's thinking. *Brit. J. Psychol.*, 1930, **20,** 354–361.

HEIDBREDER, E. Reasons used in solving problems. *J. exp. Psychol.*, 1927, **10,** 397–414.

HEIDBREDER, E. Problem solving in children and adults. *J. genet. Psychol.*, 1928, **35,** 522–545.

HEIDBREDER, E. The attainment of concepts: III. The process. *J. Psychol.*, 1947, **24,** 93–138.

HEIDBREDER, E. Experiments by Dattman and Israel on the attainment of concepts. *J. Psychol.*, 1952, **34,** 115–136.

HOFFMAN, H. N. A study in an aspect of concept formation, with subnormal, average, and superior adolescents. *Genet. psychol. Monogr.*, 1955, **52,** 191–239.

HULL, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28**(1, Whole No. 123).

KENDLER, T. S., & KENDLER, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58,** 56–60.

KENDLER, T. S., KENDLER, H. H., & WELLS, D. Reversal and nonreversal shifts in nursery school children. *J. comp. physiol. Psychol.*, 1960, **53,** 83–88.

KLINGBERG, G. The distinction between living and not living among 7–10 year old children, with some remarks concerning the so-called animism controversy. *J. genet. Psychol.*, 1957, **90,** 227–239.

LEVY, N. M., & CUDDY, J. M. Concept learning in the educationally retarded child of normal intelligence. *J. consult. Psychol.*, 1956, **20,** 445–448.

LONG, L. Conceptual relationships in children: The concept of roundness. *J. genet. Psychol.*, 1940, **57,** 289–315.

LONG, L., & WELCH, L. Reasoning ability in young children. *J. Psychol.*, 1941, **12,** 21–44.

OSLER, S. F., & TRAUTMAN, G. E. Concept attainment: II. The effect of stimulus complexity on concept attainment at two levels of intelligence. *J. exp. Psychol.*, 1961, **62,** 9–13.

PIAGET, J. *The child's conception of the world.* New York: Harcourt, Brace, 1929.

PIAGET, J. *The child's conception of physical causality.* New York: Harcourt, Brace, 1930.

PLENDERLEITH, M. Discrimination learning and discrimination reversal learning in normal and feebleminded children. *J. genet. Psychol.*, 1956, **88,** 107–112.

RUSSELL, R. W. Studies in animism: II. The development of animism. *J. genet. Psychol.*, 1940, **56,** 353–366.

RUSSELL, R. W. Studies in animism: V. Animism in older children. *J. genet. Psychol.*, 1942, **60,** 329–335.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

STEVENSON, H. W., & MCBEE, G. The learning of object and pattern discriminations by children. *J. comp. physiol. Psychol.*, 1958, **51,** 752–754.

STUART, H. N. A study of sensorimotor and conceptual thinking in children between the ages of nine and eighteen. *J. exp. Educ.*, 1935, **4,** 147–153.

THOMPSON, J. The ability of children of different grade levels to generalize on sorting tests. *J. Psychol.*, 1941, **11,** 119–126.

VINACKE, W. E. *The psychology of thinking.* New York: McGraw-Hill, 1952.

WELCH, L. The genetic development of the associational structures of abstract thinking. *J. genet. Psychol.*, 1940, **56,** 175–206.

WELCH, L., & LONG, L. The higher structural phases of concept formation of children. *J. Psychol.*, 1940, **9,** 59–95.

WELCH, L., & LONG, L. Methods used by children in solving inductive reasoning problems. *J. Psychol.*, 1942, **14,** 269–275.

WELCH, L., & LONG, L. Comparisons of the reasoning ability of two age groups. *J. genet. Psychol.*, 1943, **62,** 63–76.

(Early publication received
December 28, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 9-13

CONCEPT ATTAINMENT: II. EFFECT OF STIMULUS COMPLEXITY UPON CONCEPT ATTAINMENT AT TWO LEVELS OF INTELLIGENCE[1]

SONIA F. OSLER AND GRACE E. TRAUTMAN

Goucher College

Department of Pediatrics, Johns Hopkins University

Recent experimental work has lent support to the notion that discrimination learning and concept attainment by children may be achieved either by mediational symbolic processes or simple S-R associations. The experimental evidence also points to an association between the specific learning process characteristic of *S* and his intelligence. In a study of discrimination learning Kendler and Kendler (1959) found that fast learners achieved a reversal shift more readily than a nonreversal shift, while the opposite was true of slow learners. From this result it was inferred that fast learners utilized mediators in the original discrimination, while the slow learners achieved the discrimination through an S-R association.

In a second study involving concept attainment by induction (Osler & Fivel, 1961) *Ss* were divided into gradual and sudden learners on the basis of the learning curves they generated. It was then found that in the group of superior intelligence there were more sudden learners than gradual learners, while the reverse was true in the group of normal intelligence. On the assumption that the gradual learners build up S-R associations and that sudden learners test successive hypotheses, it is concluded that the process mediating concept attainment is a function of intelligence.

The object of the present investigation was to provide an experimental means for testing the inferred relationship between intelligence and the specific learning mechanism involved in concept attainment. If hypothesis testing is more frequent among superior than normal *Ss*, it should be possible to influence the performance of the superior group by varying the number of irrelevant dimensions, on which hypotheses can be based, in the concept exemplars. For *Ss* of normal intelligence, who tend to achieve solution by the gradual building up of an S-R association, no systematic relation between the number of stimulus dimensions and speed of solution is anticipated.

The specific prediction, therefore, is that the performance measures in concept attainment will show a sig-

[1] This investigation was supported by a grant from the National Institute of Mental Health, Psychopharmacology Service Center, made to Leon Eisenberg of Johns Hopkins University, and the senior author (Grant MY-2583).

The writers wish to express their gratitude to Mary A. Adams, Assistant Superintendent, Elementary Education, Lorne H. Woollatt, Former Director of Bureau of Research, Vernon S. Vavrina, Director, Junior and Senior High Schools, and to Arthur Lichtenstein, Director of Special Services for Pupils, Baltimore City Department of Education; and to B. Melvin Cole, Director, Elementary Education, Norris A. King, Director, Secondary Education, and Charles J. Leiman, Supervisor, Psychological Services, Baltimore County Department of Education; and to the principals and their staffs who cooperated in the conduct of the study.

They also wish to thank Nancy A. Freydberg, Sandra J. Weiss, and Frances Odette for testing *Ss* and for many suggestions concerning the conduct of the study.

nificant interaction between level of intelligence and the number of irrelevant dimensions in the concept exemplars.

METHOD

Stimuli.—One concept was studied: the number *two* represented by two types of exemplars. The first method of representation consisted of two solid black circles, $\frac{1}{4}$ in. in diameter, placed in random fashion on a 5×5 grid on white cards $2\frac{3}{4}$ in. square. The complete set consisted of 150 different patterns; these were called the *formal two* (FT) stimuli. The negative stimuli in this set consisted of the same black circles in numbers varying from one to five, excluding two, placed in unduplicated random patterns.

The second method of representation was by means of two identical pictures of common objects pasted on cards identical to those used in the FT set. The pictures, which were in a variety of colors and sizes, were cut out from children's books and were different for each of the 150 cards. This set of stimuli was called the *object two* (OT). The negative stimuli for this set consisted of the same types of pictures in numbers other than two. An effort was made to have the negative stimuli match their positive mates in general area, brightness, and color combinations. Figure 1 contains examples of the two types of stimuli; but it should be remembered that the OT stimuli were actually multicolored. Altogether there were 600 different stimulus cards.

The reason for choosing naturalistic stimuli, rather than attempting to increase the complexity of the FT stimuli by introducing the variables of color or shape, was to create very large differences in the number of irrelevant dimensions in the concept exemplars without increasing their complexity to a point which gave the children of superior intelligence or older age an advantage in familiarity or in the availability of verbal labels. The use of common objects, pretested for familiarity, permitted the introduction of an exceedingly high number of attributes without the complications just enumerated. The weakness of this method of stimulus construction is that there is no control of the amount of irrelevant information; but for the purpose of this experiment this does not seem to be crucial, as long as it was reasonable to assume a *difference* in the amount of irrelevant information.

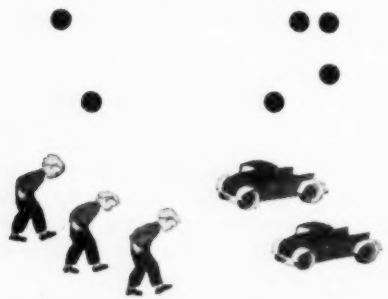

FIG. 1. Samples of the stimulus pairs used in the Formal Two (FT) and the Object Two (OT) concepts, respectively.

Prior to the actual investigation the stimulus material representing the object two was shown to 10 children in each age and intelligence group to be tested. These children were asked to identify the pictures and those which were not properly identified were discarded. The children used in this pretest did not serve as *S*s in the experiment. The familiarity of the material was checked further by asking each *S* to identify 10 pairs of stimuli upon completion of the concept learning task. Different sets of 10 were used for each *S*. These data confirmed the fact that the pictures were entirely familiar to all *S*s.

The rationale for the design of the stimuli was that the OT exemplars would suggest many more hypotheses than the FT exemplars because of the diversity in the shape, size, color, and content of the different objects represented. Preliminary work had shown that children gave three times as many different reasons for their choice of the OT as for the FT stimuli. For this reason it was anticipated that *S*s of superior intelligence would find the OT stimuli more difficult than the FT stimuli, while this should not hold for *S*s of normal intelligence.

Subjects.—The *S*s consisted of 120 school children attending elementary and junior high schools of Baltimore City and County, equally divided among 6, 10, and 14-yr. groups. Each *S* was within 6 months of the age specified. Half of the *S*s within each age level had a WISC IQ of 90 to 109 (the mean IQ of the total normal intelligence group was 101.3, $SD = 5.73$) and half had an IQ of 110 or above (the mean IQ of the total high intelligence group was 119.7, $SD = 7.46$).

Apparatus and procedure.—The apparatus and procedure have been completely described by Osler and Fivel (1961). Each *S*, who worked individually, was asked to choose the correct one of two stimuli. Reinforce-

TABLE 1

MEAN NUMBER OF ERRORS PER GROUP FOR REACHING CRITERION OF 10 CONSECUTIVE CORRECT RESPONSES

Stimuli	6-Yr. *Ss*				10-Yr. *Ss*				14-Yr. *Ss*				All Ages	
	Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ	High IQ
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	Mean
Object Two	48.6	29.5	55.1	25.3	42.1	23.4	49.1	27.2	18.2	20.2	19.3	26.3	36.3	41.2
Formal Two	48.0	28.9	36.6	16.1	47.6	24.0	23.6	29.0	15.5	19.3	9.7	12.7	37.0	23.3

Note.—Each mean is based on 10 *Ss*, total $N = 120$.

ment which followed the choice of the concept exemplar consisted of marbles to be exchanged at the end of the experiment for a prize of *S*'s selection. The *S* worked until he gave 10 consecutive correct responses, which was the criterion of success. The maximum number of trials was 150 and any *S* who did not achieve criterion performance at that point was considered to have failed. A second session, held within two weeks of the first, was devoted to the administration of the WISC.

Experimental design.—A $3 \times 2 \times 2$ factorial design was used, with three ages constituting one variable, the two methods of concept representation the second variable, and the two intelligence levels the third variable. This resulted in 12 groups with 10 *Ss* in each group, consisting of approximately the same number of boys and girls.

RESULTS AND DISCUSSION

The results were analyzed in terms of two measures of performance: (*a*) the number of errors per group, and (*b*) the number of *Ss* who attained the criterion of success. Table 1 shows the mean number of errors for each of the 12 experimental groups with the corresponding *SD*s. An analysis of variance of these data, shown in Table 2, indicates that, of the three independent variables, only age produced significant differences; neither intelligence nor mode of concept representation as main effects produced significant *F* ratios. However, as predicted, the interaction between intelligence and method of concept representation turned out to be significant at the 5% level. A breakdown of the actual data shows that the interaction is due to the fact that *Ss* of superior intelligence found the OT exemplars more difficult than the FT exemplars ($F = 7.739$, $df = 1/54$, $P < .01$); while *Ss* of normal intelligence found both types of stimuli of equal difficulty. In fact, with the OT stimuli the superior *Ss* lost all advantage of high intelligence, an advantage which was evident with the FT stimuli and also in the previous concept studies by Osler and Fivel (1961), utilizing a similar procedure.

The data on frequency of *Ss* who achieved the criterion of success are shown in Table 3. A Chi square

TABLE 2

ANALYSIS OF VARIANCE OF ERRORS

Source	*df*	*MS*	*F*
Age (A)	2	10,994.28	17.09**
Intelligence (I)	1	589.63	
Stimuli (S)	1	2,201.63	3.42
A × I	2	124.06	
A × S	2	44.31	
S × I	1	2,594.70	4.03*
A × S × I	2	363.93	
Error (Within)	108	643.45	
Total	119		

* $P < .05$.
** $P < .001$.

TABLE 3

NUMBER OF *Ss* PER GROUP WHO REACHED CRITERION PERFORMANCE

Stimuli	6-Yr. *Ss*		10-Yr. *Ss*		14-Yr. *Ss*		All Ages	
	Normal IQ	High IQ	Normal IQ	High IQ	Normal IQ	High IQ	Normal IQ	High IQ
Object Two	4	3	6	4	9	7	19	14
Formal Two	5	8	6	8	9	10	20	26

analysis shows that age ($\chi^2 = 15.60$, $df = 2$, $P < .001$) and concept representation ($\chi^2 = 6.26, df = 1, P < .02$) produced significant differences, while intelligence ($\chi^2 = .036$) did not. Because of the significant interaction in the error scores between intelligence and mode of concept representation, the frequency data were analyzed separately for each intelligence group. This analysis showed that for the normal *Ss* concept representation made no difference ($\chi^2 = .074$), while for the superior *Ss* the OT stimuli were more difficult than the FT stimuli ($\chi^2 = 10.80$, $df = 1$, $P < .01$). The results of the analysis of the frequency data are in complete agreement with those obtained from the error scores.

It has been demonstrated that the effect of increasing the number of irrelevant stimulus dimensions was to slow down *Ss* of superior intelligence without affecting those of normal intelligence. These findings are consistent with the interpretation that in concept attainment high intelligence is associated with hypothesis testing, while normal intelligence is characterized by associative learning.

In the experiment by Kendler and Kendler (1959) it was found that the use of mediating responses speeded up learning. Is the present finding of no difference between normal and superior *Ss* with the OT stimuli inconsistent with their finding? Any seeming inconsistency is easily dispelled by pointing out that where the number of irrelevant dimensions is small, as is the case with the FT stimuli, the present results are in complete accord with those reported by Kendler and Kendler, who also used simple stimuli. The more complex OT stimuli present a different problem, as is evident from the previous discussion.

Bourne and Restle (1959) recently suggested an inverse relationship between the number of irrelevant dimensions and the speed of concept identification, without restrictions as to *S* variables. They attempted a validation of this theoretical formulation with data from several studies, all of which used college students as *Ss*. Since these *Ss* are of superior intelligence, the findings are entirely consistent with those obtained in the present study for the high intelligence group.

But what about the group of normal intelligence for which the relationship proposed by Bourne and Restle did not hold? In casting about for a possible explanation one might ask whether the confounding of complexity with thematic quality of the stimuli used in the present investigation might not have altered the experimental situation. Theoretically it may have; but it is not immediately apparent why the effect of thematic stimuli should have been different for the two intelligence groups. Bruner, Goodnow, and Austin (1956, p. 108), who investigated selection strategies in concept attainment using abstract and thematic material, report that thematic material retarded the performance of their *Ss* by generating more incorrect hypotheses. Their *Ss*, who were Harvard undergraduates, are comparable to the superior group of the present study. Again, their results remain consistent with those reported here; but still this throws no additional light on the findings

with the normals. It may be that either Bourne and Restle should not extrapolate from their data obtained with superior *Ss* to those of normal intelligence; or that there is an interaction effect with the age variable whereby additional dimensions produce effects upon adults but not upon children of normal intelligence.

Summary

The aim of this investigation was to provide a means of testing a previous interpretation that *Ss* of superior intelligence attain concepts by testing hypotheses while *Ss* of normal intelligence attain concepts through S-R associative learning. One hundred and twenty *Ss*, 60 at each intelligence level, were tested on a concept attainment task. Half the *Ss* worked with simple stimuli and half with complex stimuli. In each case the concept to be attained was the same, that of the number two. Because the complex stimuli would generate more hypotheses than the simple ones, it was predicted that they would slow down the superior *Ss*, but not affect the normal *Ss*. On two measures of performance, number of errors to criterion and frequency of *Ss* attaining criterion, the prediction was confirmed.

References

Bourne, L. E., Jr., & Restle, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Kendler, T. S., & Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

Osler, S. F., & Fivel, M. W. Concept attainment: I. Effect of age and intelligence in concept attainment by induction. *J. exp. Psychol.*, 1961, **62**, 1–8.

(Early publication received December 28, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 14–23

FAMILIARIZATION (*n*) AS A STIMULUS FACTOR IN PAIRED-ASSOCIATE VERBAL LEARNING[1]

DONALD R. GANNON AND CLYDE E. NOBLE

Louisiana State University *Montana State University*

The Law of Acquaintance is currently enjoying a scientific revival among human learning psychologists. According to Robinson: "the act of reading or reciting . . . may . . . influence the facility with which that act enters into new associations" (1932, p. 118). Despite its premature announcement nearly thirty years ago, very little evidence favoring this principle was available until the past decade when the quantitative effects of frequency of stimulation (*n*) upon recognition behavior and familiarity judgments were discovered (Arnoult, 1956; Noble, 1954, 1960; Solomon & Postman, 1952; Underwood & Schulz, 1960, Exp. I). In general, the facilitation functions for a variety of materials are negatively accelerated in the range $n=0$ to 40, with returns diminishing rapidly around $n = 20$. Serial learning proficiency (R%) is similarly affected by *n* (Hovland & Kurtz, 1952; Noble, 1955; Riley & Phillips, 1959; Underwood & Schulz, 1960, Exp. III), but paired-associate learning presents a maze of contradictions.

One of G. E. Müller's students (Winzen, 1921) made an early attempt to determine whether familiarization is more influential in the stimulus (S) or in the response (R) position. Learning pronounceable syllables which they had seen either 0 or 20 times the day before, *S*s practiced by the method of right associates (Treffermethode) at a presentation rate of about one S-R pair every 9 sec. Winzen reported that the combination S_{20}-R_0 produced more correct responses than the combination S_0-R_{20}. Unfortunately, the control pairs S_0-R_0 and S_{20}-R_{20} were not included in the experimental design, nor was the statistical treatment adequate by modern standards to permit any definite conclusions. Two recent studies (Cieutat, 1960; Morikawa, 1959, Exp. II) lend partial support to Winzen's observations, but other investigators have either found null results (Bailey & Jeffrey, 1958; Mandler & Campbell, 1957; Waters, 1939) or contradictory evidence (Sheffield, 1946; Underwood & Schulz, 1960; Weiss, 1958). These experiments will be discussed later.

The present study was designed to evaluate an hypothesis based on data pertaining to *meaningfulness* (Noble, 1952a) available in 1957. Two independent experiments by Cieutat, Stockwell, and Noble (1958), confirmed since by others (Cieutat, 1959; Hunt, 1959; Morikawa, 1959, Exp. I), showed significant effects of meaningfulness on both the S term and the R term but a greater influence

[1] This paper is based upon a master's thesis performed by DRG under the guidance of CEN. An abridged report was made to the Southeastern Psychological Association, April 1958. We are grateful to Blaine L. Baker, Victor J. Cieutat, and Charles H. Koski for checking the statistical analyses; to Irmgard Taylor for assistance in translation of the Winzen article; and to Margaret B. Koch, Donald J. Lewis, and Ruth H. Preston for helpful suggestions.

Preparation of this report was facilitated by a contract between the Office of Naval Research and Montana State University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

on R than on S. Mindful of the serial learning data, they reasoned that if differences in the learning of paired associates varying in meaningfulness are due primarily to the positive transfer effects of *n*, an appropriate experiment

> might reveal that meaningfulness is a secondary phenomenon derivable from the same basic variable that controls all learning. This coordination would be demonstrated neatly if one could produce similar acquisition curves by manipulating either (frequency or meaningfulness). The empirical question [they concluded] is whether familiarization would affect the S and R terms differentially (1958, p. 201).

Since it is known that R% is a function of meaningfulness in serial and paired-associate learning (Noble, 1952b; Noble & McNeely, 1957); that R% is a function of familiarization in serial learning; and that the meaningfulness of the R term in paired associates contributes more to learning proficiency than the S term, it follows that if familiarization is the *sole* basis of the meaningfulness effect, then familiarization should facilitate the learning of paired associates in a similar manner. In symbolic terms, the following three implications were drawn by analogy to the two experiments by Cieutat, Stockwell, and Noble (1958): (*a*) $(S_{20}\text{-}R_{20}) > (S_0\text{-}R_{20}) > (S_{20}\text{-}R_0) > (S_0\text{-}R_0)$. (*b*) All main effects due to the stimulus term (S), the response term (R), and number of practice trials (N) should be significant. (*c*) Given enough practice, there ought to be significant two-factor interactions ($S \times R$, $S \times N$, $R \times N$) and a three-factor interaction ($S \times R \times N$).

Should the main effects and interactions turn out radically different from the predictions above, it would incline one toward the alternative hypothesis that familiarization (*n*) cannot be used to explain all of the variance in R% scores thought to be due to meaningfulness (*m*), and that there must be more to meaningfulness than sheer frequency of experience, e.g., number of associations (Noble, 1961; Parker & Noble, 1960). The latter is the view of *m* originally espoused by the second author (Noble, 1952a); hence the present investigation is at once an indirect test of the associationistic concept of meaningfulness and a direct test of Winzen's stimulus-familiarity hypothesis.

Method

Design considerations.—The experiment consisted of two phases: familiarization and learning. In the familiarization phase *Ss* received either 0 or 20 independent exposures (*n*) of items of low meaningfulness (*m*) and familiarity (*f*). The "words" later to be designated as S and R terms were never paired; each was presented as a single stimulus to which *S* made the identifying response of pronunciation. This insured stimulus reception as well as exercise of the response. The learning phase was designed to determine whether *n* facilitated learning, and if so, whether familiarization differentially influenced the relative contributions of S and R to the learning of five pairs of the previously isolated items. Four experimental groups were formed by all combinations of locus (S vs. R) and amount (0 vs. 20) of familiarization, while a control group received no familiarization training. All *Ss* practiced the same five S-R pairs during the learning phase, so any observed differences in R% scores among the experimental groups would have to be attributed to effects of the familiarization phase.

Several methodological points must be considered. First, high *m* theoretically requires high *n*, but not conversely (Noble, 1953); familiarization effects in this study may not, therefore, be explained as changes in meaningfulness. One experiment consistent with this view (Riley & Phillips, 1959) found that although serial learning of CVC material was accelerated by prior familiarization, the *m* values of the CVCs were unaffected. Second, it is important in the present investigation to use independent groups of naive *Ss* in a constant-trials design. The purpose of this is to study acquisition phenomena including interactions, to avoid interference effects, to permit

unbiased tests of S-R interaction, and to provide controls for generalized familiarization influences. Finally, in the interest of standardization, the familiarization and transfer operations are similar to those employed in earlier studies in this series.

Apparatus.—Instrumentation consisted of a Viewlex Viewmatic automatic slide projector (No. RCP-1) for the familiarization phase and a Patterson memory drum (No. 1-A) for the learning phase. The projector was equipped with a circular slide cartridge having a capacity of 30 2 × 2 in. slides. Four cartridges were required, each of which contained the familiarization items for a single experimental group. The memory drum was fitted with a white tape containing the paired associates typed in capital letters. A sliding shutter over the aperture permitted *S* to view either the left or the right side of the drum. The projection and tachistoscopic functions were controlled by two Hunter electronic timers set to expose each slide for 1 sec., and to provide a 1.4-sec. interstimulus period.[2] This gave a presentation rate of one slide every 2.4 sec. in the familiarization phase. For the learning phase, the intertrial interval was 8 sec., and the S item appeared for 2 sec. followed by the S and R items together for 2 sec.

Learning materials.—Fifteen dissyllabic "words," mostly paralogs, were used for familiarization training. Ten of these (relevant items) also appeared as the S-R elements of the paired-associate lists in the learning phase, while the remaining five (irrelevant items) were used only as fillers in the familiarization phase in order to equate the total number of exposures for *Ss* in the four experimental conditions. All items were chosen from the low end of the *m* and *f* scales (Noble, 1953) so that the mean median *m* value was .30, and the mean *f* value was .41. Low scale values were used in order to reduce the number of possible S-R connections formed outside the experimental situation and to insure comparability with the previous serial learning study (Noble, 1955). The four familiarization lists are reproduced in Table 1.

Subjects.—The *Ss* were 120 male and female college students obtained on a voluntary basis at Louisiana State University in 1957. Ages ranged from 17 to 43 yr. (mean = 19.3 yr.), and all *Ss* were naive with respect to paired-associate learning. Twenty-four *Ss* were assigned to each of the four experimental groups in a counterbalanced order of arrival at the laboratory. An additional 24 *Ss* were used as a special control group which received no familiarization training. The *E* engaged these *Ss* in conversation for an amount of time equivalent to the familiarization phase, then proceeded to the learning phase.

Procedure.—Four experimental groups were formed in terms of locus and amount of familiarization, as follows: S_{20}-R_{20}, S_0-R_{20}, S_{20}-R_0, S_0-R_0. In the familiarization phase, *S* was instructed to pronounce each dissyllable aloud as it was projected on the screen. The *Ss* were not informed of the subsequent learning phase. As shown above in Table 1, the relevant items were separated by irrelevant items in order to reduce the formation of any initial S-R connections before the learning phase. The ratios of relevant to irrelevant items in the S_{20}-R_{20}, S_0-R_{20}, S_{20}-R_0, and S_0-R_0 groups, respectively, were 20/10, 10/20, 10/20, and 0/30. Since each relevant item was exposed

TABLE 1

FAMILIARIZATION LISTS FOR THE FOUR EXPERIMENTAL GROUPS APPEARING IN EACH PROJECTOR CARTRIDGE

	S_{20}-R_{20}	S_0-R_{20}	S_{20}-R_0	S_0-R_0
I	DELPIN	DELPIN	DELPIN	DELPIN
	BYSSUS	BYSSUS	BYSSUS	BYSSUS
R	ULNA	ZUMAP	ULNA	JETSAM
	TAROP	JETSAM	TAROP	ZUMAP
	GOKEM	DELPIN	GOKEM	BYSSUS
	NOSTAW	QUIPSON	NOSTAW	QUIPSON
	SAGROLE	BYSSUS	SAGROLE	ZUMAP
	NOSTAW	ZUMAP	NOSTAW	DELPIN
	GOKEM	DELPIN	GOKEM	QUIPSON
	TAROP	JETSAM	TAROP	JETSAM
	ULNA	QUIPSON	ULNA	BYSSUS
	SAGROLE	BYSSUS	SAGROLE	DELPIN
I	JETSAM	JETSAM	JETSAM	JETSAM
	ZUMAP	ZUMAP	ZUMAP	ZUMAP
	QUIPSON	QUIPSON	QUIPSON	QUIPSON
	BYSSUS	BYSSUS	BYSSUS	BYSSUS
	ZUMAP	ZUMAP	ZUMAP	ZUMAP
	DELPIN	DELPIN	DELPIN	DELPIN
	VOLVAP	VOLVAP	ZUMAP	QUIPSON
	XYLEM	XYLEM	JETSAM	DELPIN
	MEARDON	MEARDON	DELPIN	ZUMAP
	KUPOD	KUPOD	QUIPSON	JETSAM
R	LATUK	LATUK	BYSSUS	DELPIN
	KUPOD	KUPOD	ZUMAP	BYSSUS
	MEARDON	MEARDON	DELPIN	QUIPSON
	XYLEM	XYLEM	JETSAM	JETSAM
	VOLVAP	VOLVAP	QUIPSON	BYSSUS
	LATUK	LATUK	BYSSUS	ZUMAP
I	QUIPSON	QUIPSON	QUIPSON	QUIPSON
	JETSAM	JETSAM	JETSAM	JETSAM
I	5@2 ea.	5@4 ea.	5@4 ea.	5@6 ea.
R	10@2 ea.	5@2 ea.	5@2 ea.	none
Σ	30	30	30	30

Note.—R = relevant items, I = irrelevant items.

[2] The present exposure duration is from 1.3 to 2.0 sec. briefer than those used in earlier studies (Noble, 1954, 1955), but the interstimulus period is the same.

twice in each cartridge, 10 revolutions were required to obtain 20 exposures. There were 300 exposures in all.

In the learning phase, *S* was instructed to anticipate the R term after the first trial. As is standard in this series, both S and R were pronounced and the correction method was used. Seventeen learning trials were given, with the two groups of five relevant items constituting the S and R terms. Responses were scored as correct (R+) or erroneous (R−). The criteria for an erroneous response were as follows: (*a*) omissions, (*b*) late responses, (*c*) mispronounced responses, (*d*) the same response repeated three times in succession during a single trial. The *Ss* had to maintain a uniform pronunciation of the paralogs in order to receive R+ scores by the third error criterion. To neutralize accidental order effects, two lists of S-R pairs were formed. The Forward list, mounted on the left side of the memory drum, consisted of pairs formed from the 10 relevant items. The Reverse list, mounted on the right side of the drum, transposed these items so that S_1 became R_1, S_2 became R_2, etc. The *E* constructed four random trial sequences to minimize serial learning effects. These learning lists are reproduced in Table 2. There were three Forward and three Reverse *Ss* assigned to each of the four trial sequences, making a total of 24 *Ss* per condition. The familiarization and learning phases required 12 min. each, separated by a 1-min. rest interval during which *E* engaged *S* in conversation. Total time, including instructions, was approximately 27 min. for all *Ss*.

Instructions.—The familiarization instructions were given verbatim as follows:

> This is an experiment on verbal learning. We are interested in the general learning process common to all people and are not testing your intelligence or personality. In the first part of the experiment you will be shown slides to get you acquainted with the procedure. With this projector, I am going to flash on the screen some two-syllable words similar to actual words. Probably you have never seen any of them before so there is no standard or correct pronunciation. All you have to do is say each word out loud before it leaves the screen. Whichever way you pronounce each word when we start is all right, but try to say it the same way each time the word comes up. Do *not* try to memorize the order of the words because this will change constantly. Please say each word loudly and clearly enough so I can hear and understand you. In this way I will be sure to give you credit for a response. Try to say the word before it leaves the screen, since the words come in rapid order. Are there any questions? Now remember, if you see "dinner," say "dinner." Ready?

The learning instructions were given verbatim as follows:

TABLE 2

PAIRED-ASSOCIATE LISTS AND TRIAL SEQUENCES FOR ALL GROUPS DURING THE LEARNING PHASE

Trial Sequences	Learning Lists			
	Forward		Reverse	
I	TAROP	XYLEM	XYLEM	TAROP
	SAGROLE	LATUK	LATUK	SAGROLE
	GOKEM	MEARDON	MEARDON	GOKEM
	ULNA	VOLVAP	VOLVAP	ULNA
	NOSTAW	KUPOD	KUPOD	NOSTAW
II	ULNA	VOLVAP	VOLVAP	ULNA
	GOKEM	MEARDON	MEARDON	GOKEM
	TAROP	XYLEM	XYLEM	TAROP
	NOSTAW	KUPOD	KUPOD	NOSTAW
	SAGROLE	LATUK	LATUK	SAGROLE
III	GOKEM	MEARDON	MEARDON	GOKEM
	NOSTAW	KUPOD	KUPOD	NOSTAW
	ULNA	VOLVAP	VOLVAP	ULNA
	SAGROLE	LATUK	LATUK	SAGROLE
	TAROP	XYLEM	XYLEM	TAROP
IV	NOSTAW	KUPOD	KUPOD	NOSTAW
	ULNA	VOLVAP	VOLVAP	ULNA
	SAGROLE	LATUK	LATUK	SAGROLE
	TAROP	XYLEM	XYLEM	TAROP
	GOKEM	MEARDON	MEARDON	GOKEM

> This is the second part of the experiment. Here is a memory drum. Shortly after the apparatus starts, you will see a word in the window. You are to pronounce this word and those that follow it as you see them. First, a word will appear alone, then the shade will rise and you will see another word. You are to pronounce them both. After you have seen this entire list once, your job will be to call out each single word and then to anticipate the word with which it was paired. In other words, as you see the *single* word, you are to pronounce it, and then immediately you are to pronounce the word with which it was paired *before* the second word appears. If you fail to anticipate an item, pronounce it when it appears anyway. If you think you know what a word will be but are not sure, *guess*. It will not hurt your score any more than to say nothing, and if you get it correct it will count as a success. If you anticipate a word *incorrectly*, correct yourself as soon as the word appears. Remember to pronounce each item *aloud*. Do not try to think ahead more than one step at a time or to count, because the *order* of the pairs will change as the list is repeated. Do not try to use any special system in your learning. Simply associate each single word with the word with which it is paired. Please try to pronounce all words the same way each time they appear so that I can give you credit for a correct response. Any questions?

RESULTS

The major results of the experiment are shown in Fig. 1. Percentages of correct responses (R%) are plotted as a function of amount of practice (N) grouped in blocks of two trials for the five conditions. The influence of S-term familiarization is greater

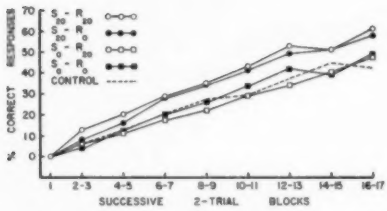

FIG. 1. Acquisition curves for a list of five paired associates as a function of practice under four different combinations of low ($n = 0$) and high ($n = 20$) frequency of stimulus (S) and response (R) term familiarization. (The Control group received no prior familiarization experience. Each curve contains 24 *Ss*.)

than that of R-term familiarization, while the effect of the latter treatment is little different from having no familiarization at all. Total R+ scores for the various groups during the 17-trial period, given in parentheses, were ordered as follows: S_{20}-R_{20} (735), S_{20}-R_0 (689), S_0-R_0 (542), Control (524), S_0-R_{20} (504). The S-term difference is (1424−1046 =) 378 units, contrasted with an R-difference of only (1239−1231 =) 8 units. An 8 × 5 Type I mixed-factorial analysis of variance (Lindquist, 1953, p. 267) was performed on the R+ scores from which Fig. 1 was constructed. There were significant main effects of Practice ($F = 157.29$; $df = 7/805$; $P < .001$) and Conditions ($F = 2.69$; $df = 4/115$; $P < .05$), but the interaction was not significant ($F < 1.00$; $df = 28/805$).

In order to test for a learning-to-learn familiarization effect, the Control group may be considered a baseline from which to measure possible generalized influences of the familiarization phase on the learning phase. A 3 × 1 simple-randomized analysis of variance (Lindquist, 1953, p. 47) was therefore applied to the overall R+ scores of the Control group and the two S_0 groups, omitting both S_{20} groups. The result gave no evidence of any nonspecific familiarization effect ($F < 1.00$; $df = 2/68$; $P > .20$), so the three lowest curves in Fig. 1 may be regarded as the same except for sampling fluctuations. There is neither a relevant response effect (S_0-R_{20} vs. S_0-R_0) nor an irrelevant stimulus-response effect (S_0-R_0 vs. Control) that can be attributed to the 300 trials of stimulation and pronunciation experienced during the familiarization phase.

To evaluate possible two-factor and three-factor interactions, a 16×2 × 2, N × S × R Type III mixed-factorial analysis of variance (Lindquist, 1953, p. 281) was performed on the individual trial R+ scores for the four experimental groups only. The summary, shown in Table 3, indicates significant main effects of Practice (N) and Stimulus Familiarization (S). Neither the Response Familiarization (R) factor nor any of the interactions were significant sources of variance.

Because R+ scores do not reveal any Practice × Conditions interactions, the facilitation due to S-term familiarization must be largely an initial effect followed by latter paral-

TABLE 3

ANALYSIS OF VARIANCE OF CORRECT RESPONSES (R+) OF THE EXPERIMENTAL GROUPS DURING TRIALS 2–17

Source	*df*	*MS*	*F*
Between *Ss*	95		
Stimulus (S)	1	93.02	7.86*
Response (R)	1	.04	.00
S × R	1	4.60	.39
Error (*b*)	92	11.83	
Within *Ss*	1440		
Trials (N)	15	57.99	84.30**
S × N	15	.70	1.03
R × N	15	.44	.63
S × R × N	15	.54	.78
Error (*w*)	1380	.69	
Total	1535		

* $P < .01$.
** $P < .001$.

lelism. This hypothesis is confirmed by comparing the average number of trials or errors (R−) preceding the first correct anticipation of each S-R pair in all groups. Most of these R− scores were omissions rather than late or overt incorrect responses, so no qualitative analysis of errors will be made. Mean R− scores required to attain this reaction-threshold criterion (averaged over *Ss* not S-R pairs) were ordered as follows: S_{20}-R_{20} (5.20), S_{20}-R_0 (5.33), Control (6.28), S_0-R_{20} (6.88), S_0-R_0 (6.95). A 5 × 1 simple-randomized analysis of variance (Lindquist, 1953, p. 47) of these data gave an F ratio of 2.79 ($df = 4/112$; $P < .05$) which is significant. Again, the major difference favors the two S_{20} groups over the S_0 groups, with the Control group in the middle. Employing t tests of the deviations of each pair of extreme groups from the Control group based on the within-groups σ^2 (Lindquist, 1953, p. 91), we find significantly more errors under the S_0 (irrelevant n) combination and fewer errors under the S_{20} (relevant n) combination ($t = 3.31$; $df = 112$; $P < .01$); however, the two component ratios are not significant individually ($t \geqslant 1.69$; $df = 112$; $P > .05$). By the second correct (reinforced) anticipation all five groups have converged ($F < 1.00$; $df = 4/110$), and the speed of error elimination is essentially constant thereafter in agreement with Fig. 1. Observed group trends in R+ scores thus probably reflect unequal levels of proficiency (number of associates attained per trial) rather than differential rates of learning (growth of habit strength).

The results lead us to state that n has a facilitative influence on paired-associate learning only when familiarized items become the S members. Prelearning familiarization with the R members fails to produce any significant result, and except for initial error effects there is no further interaction between locus of familiarization (S vs. R) and amount of practice (N). In the light of these data, a simple stimulus-frequency theory of meaningfulness (Cieutat et al., 1958; Underwood & Schulz, 1960) must be rejected as untenable. We conclude that familiarization (n) does not produce effects which are equivalent to those produced by meaningfulness (m), and that m and n are therefore independent variables affecting verbal behavior.

Discussion

Two principal questions are raised by this experiment: (*a*) Are the results comparable with those obtained by other investigators? (*b*) Assuming no artifacts, what is the most reasonable interpretation of the current data?

Comparability.—Eight laboratories have followed up Winzen's (1921) work. Negative results in the Waters (1939), Mandler and Campbell (1957), and Bailey and Jeffrey (1958) experiments seem to be contingent on the use of CVC or CCC materials and very brief amounts or durations of familiarization. Contradictory findings reported in the Sheffield (1946), Weiss (1958), and Underwood and Schulz (1960) experiments are also associated with trigram material; in addition, the types and amounts of familiarization as well as the transfer tests are different. In opposition to Winzen, the hypothesis coming from this second group of studies suggests that familiarization is effective only in the R position. Occasionally there were signs of decremental influences of S-term familiarization (Sheffield, 1946; Underwood & Schulz, 1960, Exp. I, IV), but these trends were neither significant nor consistent (Underwood & Schulz, 1960, Exp. II). Although the latter experiments contain many ingenious features, there are difficulties in evaluating the R-facilitation and S-interference phenomena because of rather complex familiar-

ization techniques, sophisticated *Ss*, and incomplete factorial designs. Statistical tests of interaction among the S, R, and N factors are frequently lacking. As Underwood and Schulz say of their Exp. I, II, and IV: "they have raised more issues than they have settled regarding the effects of stimulus familiarization" (1960, pp. 125–126).

In support of Winzen, on the other hand, are the present data together with those of Morikawa (1959, Exp. II) and Cieutat (1960). Morikawa noted definite S-facilitation in Japanese *Ss* when the R term was unfamiliar but little difference when the R term was familiar. A different procedural approach was tried by Cieutat in that pronounceable items of low *m* value were used, but he varied exposure duration (0 vs. 60 sec.) of the two terms (S vs. R) rather than frequency and did not require verbalization. Another difference is that his *Ss* served in all four combinations of duration and locus; however, both S and R were pronounced in the transfer period, as in this experiment, and the *Ss* were from the same population. Cieutat found significant effects attributable to N and to the S × R interaction, the latter being observed as a weak S-term decrement under R_0, a strong S-term increment under R_{60}, and an overall main effect favoring stimulus familiarization. (Referring to his Table 2, mean R+ scores for the S_0 and S_{60} conditions were 2.31 and 2.50, respectively, whereas means for R_0 and R_{60} were both exactly 2.41 units.) Sustained viewing, possible intertreatment interferences, and failure to require articulation during the familiarization period may account for the slight (statistically marginal) inhibitory S-effect under irrelevant R-term familiarization; the absence of a reference level provided by an unfamiliarized control group precludes firm inferences. This peculiarity is also characteristic of the Sheffield and Underwood-Schulz experiments. In one crucial respect, however, Cieutat's data are quite consistent with Winzen's and ours; namely, the significant superiority of his S_{60}-R_0 treatment combination over that of S_0-R_{60}.

The use of confounded and incomplete designs, unrelated or inadequate control procedures, *Ss* of heterogeneous experience, diverse types or amounts of familiarization, different learning materials, and varied methods of testing for transfer renders impossible any comparative evaluation of the present findings in the same context with the above studies. About all that can be done from a constructive point of view is to urge that the reader be cautious in trying to resolve these conflicting reports, and to judge each set of experimental data in the light of its own methodology. With the reminder that our procedure was standard in this research program (e.g., specific familiarization technique), and that it was designed to meet certain boundary conditions (e.g., two-term articulation during learning), we turn now to a tentative explanation of the results.

Interpretation.—The data permit three unequivocal statements: (*a*) relevant S-term familiarization produced significant superiority in total R+ scores whereas relevant R-term familiarization had no influence; (*b*) when compared with the unfamiliarized Control condition, initial R− scores increased after irrelevant S-term familiarization and decreased after relevant S-term familiarization; (*c*) acquisition rates of correct responses and extinction rates of incorrect responses both became independent of original familiarization following the first reinforcement in the list.

Recalling the introductory discussion, a two-factor hypothesis may be suggested to account for the differential effects of *m* and *n* in paired-associate learning under these conditions. The evidence indicates that *n* is sufficient to influence behavior via the S term alone (stimulus predifferentiation) whereas *m* is *jointly* necessary to affect behavior via the R term (response patterning). One possible explanation of S-term facilitation, considered earlier by the second author in connection with the complex familiarization effects in serial learning (where the "perceptual" and "motor" aspects of verbal learning are confounded), is provided by the principle of stimulus

constancy (Noble, 1955, p. 336). The core assumption, derived from the conditioning literature, is that repeated exposure-articulation sequences during the familiarization phase will attenuate variability in *S*'s perceptual (e.g., orientation, fixation) and identifying (e.g., pronunciation) responses to the paired-associate S terms when the learning phase is encountered. Assuming further that the stability of these perceptual and identifying responses grows in some negatively accelerated fashion with increasing *n* (like habit strength or familiarity), their facilitative influences ought to be maximal early in the learning phase then subside later as paired-associate training progresses. Although less parsimonious than a simple frequency explanation, this notion helps to understand the consistent (but weak) S-effects of *m* which have been observed in a number of studies (Cieutat, 1959; Cieutat et al., 1958, Exp. I, II; Hunt, 1959; Morikawa, 1959, Exp. II; Underwood & Schulz, 1960).[3]

Our two-factor hypothesis implies a systematic view of the verbal learning process which is somewhat at variance with that held by Underwood and Schulz (1960). Their two-stage analysis of "response-recall" followed by "associative hook-up" (reminiscent of Thorndikean selecting and connecting) postulates that frequency alone (e.g., *n* without reinforcement) is important for the first stage of learning, the second being related to meaningfulness (or pronunciability). So far we might agree, but they proceed from their "spew" hypothesis, which is based on free-recall phenomena, yet considered "not relevant for stimulus familiarization" (p. 100), to predict that (*a*) "rate of paired-associate learning should be an increasing function of amount of familiarization training given the response units of the list," and (*b*) that "response familiarization should facilitate learning considerably more than stimulus familiarization" (pp. 100–101). The emphatic lack of any R-term effect or of an R > S relation in the present data is clearly embarrassing to the Underwood-Schulz hypothesis.

But is our evidence relevant? A sophisticated critic might object that this investigation (using pronounceable words and paralogs) is not a fair test of the "spew" hypothesis because it fails to satisfy some of the Underwood-Schulz boundary conditions (e.g., CVCs and the spelling method). Unfortunately, these writers are neither very specific nor consistent about such matters (cf. the familiarization procedures, transfer tests, and *S*s in Exp. I vs. Exp. IV), but in one chapter (p. 95) they describe a hypothetical task where, even though the responses are "well integrated" (pronounceable), prior differences in *n* are definitely expected to alter the "availability" (reaction threshold?) of the R terms. The present procedure appears to conform to these general requirements; only the data are recalcitrant, paradoxically. Considering their repeated emphasis on the first stage process, incidentally, we find it curious that Underwood and Schulz are less interested in initial performance and rate measurements than in overall analyses.

A second possible objection to this experiment turns on our custom in this series of having *S*s articulate both terms in the paired-associate learning phase. It may be claimed that this requirement either (*a*) produces an unusual amount of S-term familiarization, or (*b*) shortens the effective anticipation interval for the S_0 groups as contrasted with the S_{20} groups. Briefly (aside from the standardization issue), our rejoinder to (*a*) is that we believe double articulation is necessary both to guarantee stimulus reception and to provide an unbiased test of differences between the manipulated properties of the antecedent (S) and consequent (R) events. Since all factors except temporal succession should be equated when evaluating the effects

[3] This tentative explanation does not presume to account for the initial decrements in the S_0-R_0 and S_0-R_{20} groups (relative to the Control), but it would not be inconsistent with a theory of stimulus variability to hypothesize that irrelevant familiarization in the S-position produces interference via stimulus trace generalization.

of *S*'s built-in response to an S term upon his ability to anticipate its paired R term, it seems elementary to have the former occur just as frequently as the latter during the learning phase (cf. also the rationale of the correction method on erroneous trials). This Underwood and Schulz did not do, hence it constitutes yet another difference between our procedures. Such a criticism could be turned around and directed at those who introduce biases favoring R-term familiarization.

The distributed-practice argument (*b*) is more plausible, though it loses force in view of Hunt's (1959) report of null results in a Cieutat-Stockwell-Noble (1958, Exp. I) type of experiment in which half the *S*s in each group articulated the paralogs while the other half did not. It is important to note that Hunt's material, the same dissyllables of the *m* scale used herein, can be pronounced faster than CVCs can be spelled. Of course, no one can be sure without experimenting that S-R familiarization effects do not interact with distributional factors, but the likelihood appears slim in the present case because *S*s in all groups except the Control were trained for 300 trials to pace their pronunciation responses at a 1-sec. rate during the familiarization phase. Furthermore, the S_0-R_0 group is slightly superior to both the Control and the S_0-R_{20} groups on R+ performance throughout the learning phase. The same phenomenon occurs in Cieutat's (1960) study. Both "spew" and distributional hypotheses imply the exact opposite of these observations. Appropriate experiments to settle this question will occur readily to most readers; meanwhile, the second author has some research in progress which indicates that type and tate of *n* are also relevant. The S-articulation factor is yet to be explored.

Our best judgment at the present writing is that no serious artifacts or procedural errors are contaminating the interpretation presented above. To repeat, however, the familiarization issue is not yet definitely settled. Much undoubtedly remains to be learned about the experimental control of *n* and the measurement of its effects.

Summary

The role of familiarization (*n*) in paired-associate verbal learning was investigated. One hundred twenty *S*s, divided into five independent groups of 24 each, were used. Four experimental groups were formed by all combinations of locus (S vs. R term) and amount (0 vs. 20 exposures) of familiarization, while a Control group received no familiarization experience. All groups then practiced a five-unit paired-associate list of dissyllables of low initial familiarity (*f*) and meaningfulness (*m*) for 17 trials.

The influence of relevant S-term familiarization was positive and significant ($P < .01$), but not that of relevant R-term familiarization. Analyses of correct responses (R+) revealed no evidence of an irrelevant (learning-to-learn) familiarization effect, nor of any interactions between S or R and amount of practice. The initial effects of familiarization on incorrect responses (R−) were also significant ($P < .05$), causing decrements in the (irrelevant) S_0 groups and increments in the (relevant) S_{20} groups when compared with nonfamiliarized controls. After the first correct (reinforced) anticipation, acquisition rates of R+ and extinction rates of R− were uniform for all groups.

It was concluded that familiarization (*n*) and meaningfulness (*m*) operations do not produce equivalent effects on verbal behavior under these conditions. Contradictory results from other laboratories were discussed, and a tentative two-factor hypothesis was presented as an attempted synthesis of the differential roles played by *m* and *n* in paired-associate learning.

References

Arnoult, M. D. Familiarity and recognition of nonsense shapes. *J. exp. Psychol.*, 1956, **51**, 269–276.

Bailey, J. H., & Jeffrey, W. E. Response strength and association value in stimulus predifferentiation. *Psychol. Rep.*, 1958, **4**, 715–721.

Cieutat, V. J. Supplementary report: Stimulus and response meaningfulness (*m'*) in paired-associate learning by hospitalized mental patients. *J. exp. Psychol.*, 1959, **58**, 490.

Cieutat, V. J. Differential familiarity with stimulus and response in paired-associate

learning. *Percept. mot. Skills*, 1960, **11**, 269–275.

CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (m, m') in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 193–202.

HOVLAND, C. I., & KURTZ, K. Experimental studies in rote-learning theory: X. Pre-learning syllable familiarization and the length-difficulty relationship. *J. exp. Psychol.*, 1952, **44**, 31–39.

HUNT, R. G. Meaningfulness and articulation of stimulus and response in paired-associate learning and stimulus recall. *J. exp. Psychol.*, 1959, **57**, 262–267.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

MANDLER, G., & CAMPBELL, E. H. Effect of variation in associative frequency of stimulus and response members on paired-associate learning. *J. exp. Psychol.*, 1957, **54**, 269–273.

MORIKAWA, Y. Functions of stimulus and response in paired-associate verbal learning. *Psychologia*, 1959, **2**, 41–56.

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430. (a)

NOBLE, C. E. The role of stimulus meaning (m) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446; **44**, 465. (b)

NOBLE, C. E. The meaning-familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89–98.

NOBLE, C. E. The familiarity-frequency relationship. *J. exp. Psychol.*, 1954, **47**, 13–16.

NOBLE, C. E. The effect of familiarization upon serial verbal learning. *J. exp. Psychol.*, 1955, **49**, 333–338.

NOBLE, C. E. Supplementary report: Familiarity and frequency. *J. exp. Psychol.*, 1960, **59**, 432–433; **60**, 418.

NOBLE, C. E. Verbal learning and individual differences. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961, in press. Ch. 6.

NOBLE, C. E., & MCNEELY, D. A. The role of meaningfulness (m) in paired-associate verbal learning. *J. exp. Psychol.*, 1957, **53**, 16–22.

PARKER, G. V. C., & NOBLE, C. E. Effects of experimentally-produced meaningfulness (m) on paired-associate learning. *Amer. Psychologist*, 1960, **15**, 451. (Abstract)

RILEY, D. A., & PHILLIPS, L. W. The effects of syllable familiarization on rote learning, association value, and reminiscence. *J. exp. Psychol.*, 1959, **57**, 372–379.

ROBINSON, E. S. *Association theory today*. New York: Century, 1932.

SHEFFIELD, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished doctoral dissertation, Yale University, 1946.

SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–201.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Chicago: Lippincott, 1960.

WATERS, R. H. The law of acquaintance. *J. exp. Psychol.*, 1939, **24**, 180–191.

WEISS, R. L. The role of association value and experimentally produced familiarity in paired associate verbal learning. (Unpublished doctoral dissertation, University of Buffalo, 1958) *Amer. Psychologist*, 1958, **13**, 386. (Abstract)

WINZEN, K. Die Abhängigkeit der paarweisen Assoziation von der Stellung des besser haftenden Gliedes. *Z. Psychol.*, 1921, **86**, 236–252.

(Early publication received August 17, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 24–34

PAIRED-ASSOCIATES LEARNING WITH VARYING RELATIVE PERCENTAGES OF OCCURRENCE OF ALTERNATIVE RESPONSE MEMBERS[1]

ALBERT E. GOSS AND MARILYN E. SUGERMAN

University of Massachusetts

The purpose of the present experiment was to investigate learning with divergent paired-associates (PA) units: each divergent PA unit consisted of a single stimulus member and two response members, one of which occurred on $p\%$ and the other of which occurred on $(1\text{-}p)\%$ of the presentations of the stimulus member. One of the two major variables was relative percentages (RP) of occurrence of the alternative response members of each unit; the other was number of units.

Voss, Thompson, and Keegan (1959) have reported what are apparently the first experiments on acquisition with divergent units under conventional conditions of PA learning. In Exp. I and II, a different one of six RP from 50–50 to 100–0 in 10% steps was assigned in counterbalanced fashion to one of the units of a six-unit list. In Exp. III, the same PA list was used, but one of five RP from 50–50 to 90–10 was assigned to all six units. Whether the same *S*s had all or but one of the RP, percentages of anticipations of the more frequent response member for RP from 60–40 to 90–10 exceeded actual percentages of occurrence of those members. Anticipations of the alternative responses of the 50–50 RP were at or below 50%.

[1] Sponsored by the Office of Naval Research under Contract Nonr 2691(00). Susan Gregory helped to prepare materials for the experiment, Nancy Farrick assisted in the collection and analyses of data, and Alice Reilly aided in analyses of data.

The present experiment complemented and extended those of Voss, Thompson, and Keegan. First, CVC nonsense trigrams of the present experiment were more conventional in form and origin than their CVCC units from a newly developed list. Second, the one-unit and four-unit lists of the present experiment filled the gap below their list with the same RP for all six units. Third, the one-unit and four-unit lists constituted a further variable: number of divergent units in a list, each with the same RP.

Number of units was introduced because of the general importance of length of material for learning and performance. The one-unit value was selected in part because it was the minimum length. More importantly, the stimulus-response relationships of a single divergent PA unit seemed homologous to those of certain simple two-choice situations for human *S*s involving a single first stimulus followed by one or the other of two second stimuli, each of which presumably evokes a different response.[2]

[2] Simple two-choice situations which involve a single first stimulus followed only by one or the other of two second stimuli should be distinguished from those which involve a single first stimulus and not only one or the other of two second stimuli but also additional correlated stimuli (e.g., one amount of reward accompanying one of the second stimuli and another amount of reward or loss of reward accompanying the other of the second stimuli). Only findings and theory for the former situation are pertinent to learning with divergent PA units whose second stimuli are only the response members of a unit.

Accordingly, the one-unit list tested the generality of equations, such as Estes' (1959b, p. 411), whose implications for a single divergent PA unit with two response members are that percentages of anticipations of the more frequent (or the less frequent) response member will approach actual percentages of occurrence of that response member in negatively-accelerated fashion, at a rate independent of RP. Lists of n divergent PA units can be conceptually decomposed into n single units. The four-unit list, therefore, tested whether or not predictions based on equations and findings for simple two-choice situations and also for a single divergent unit could be extended without modification to lists of more than one unit.

METHOD

Experimental design.—Six relative percentages of occurrence of alternative response members of each PA unit (50–50, 60–40, 70–30, 80–20, 90–10, 100–0) were used with a four-unit list and with one-unit lists involving three conditions of presentations. (Hereafter the four-unit list and the three conditions of presentation are collectively referred to as lists, list conditions, or L.) The three conditions of presentation controlled for possible differences in learning with four-unit and one-unit lists attributable to length and variability of time intervals between successive occurrences of the same PA unit, and to the nature of the stimuli and *Ss*' activities during those intervals.

With a four-unit list, each repetition of a particular unit is, on the average, separated by occurrences of the three other units. With a one-unit list, the time interval between each occurrence of the unit might be the same as the time interval between each of the four different units of a four-unit list. If so, the time interval between the single unit would be short and constant, while the time intervals between repetitions of the same unit in a four-unit list would be longer and variable. The *one-unit* condition involved a constant 2-sec. interval between successive presentations of the paired-associates unit. The *one-unit spaced* condition involved time intervals between successive presentations of the paired-associates unit whose average length and variability of the time intervals between successive presentations of that particular unit in a four-unit list. The one-unit spaced condition did not control for differences between four-unit and one-unit spaced lists with respect to rehearsal, associations with syllables of other units, and warm up. With the *one-unit random* condition, in order to minimize rehearsal of the responses of a particular unit and to provide warm-up experiences as well as the possibility of associations between the syllables of that particular unit and those of other units, time intervals between successive occurrences of the same paired-associates unit were filled by occurrences of pairs of syllables whose pairing was random rather than constant.

Subjects.—Men and women students enrolled in the introductory course in psychology at the University of Massachusetts participated as *Ss* in partial fullfillment of course requirements. The 24 combinations of RP and L were run in eight counterbalanced cycles, with *Ss* being assigned to a particular combination in order of its occurrence and *Ss*' appearance until there were 8 *Ss* in each combination, a total of 192 *Ss*. Data from 28 other *Ss* could not be used because of *Es*' errors ($N = 8$), apparatus failures ($N = 8$), or *Ss*' inability of refusal to continue ($N = 12$). Seven of the latter 12 were distributed among three RP with the four-unit list; the remainder were distributed among the three one-unit conditions.

Lists.—The CVC trigrams of the PA units of the four-unit and one-unit lists, with the recently obtained Archer (1960) association values in parentheses, were WOM (67%) followed by TEX (100%) or LIK (91%), RUF (86%) followed by DOL (94%) or CEN (75%), FES (69%) followed by VIC (97%) or BAL (91%), and NAV (72%) followed by HOB (90%) or SUR (91%). None has more than one letter in common with any other. The stimulus member and the one response member of each PA unit of a given trial were typed in elite capitals on the same line, 3 in. apart, on white paper, then attached to white plastic cards and presented by Hunter Card Master.

(*a*) *Four-unit.* Each trial consisted of one presentation of all four of the stimulus members, each of which was accompanied by one or the other of the two response members of that unit. Orders of occurrence of the four stimulus members during each trial, regardless of occurrences of response members, were determined by selecting two groups of 10 of the 24 permutations of four events. The

10 permutations of one group differed from those of the other group. Within each of these groups of 10 permutations, occurrence of a particular permutation was randomly determined. Since 120 trials were administered, each of the two groups of 10 permutations was repeated in a sequence which was determined randomly, subject to the constraints that each group of 10 permutations should occur 6 times and that neither group should occur more than 3 times in succession.

For the 50–50 RP, WOM was followed by LIK 50% of the time and by TEX 50% of the time, and similarly for the pairs of response members for RUF, FES, and NAV.

Under each of the other relative percentages, TEX, DOL, VIC, and HOB occurred with the same larger or smaller percentage and LIK, CEN, BAL, and SUR occurred with same smaller or larger percentage, respectively. For example, for half of the *Ss* with RP of 60–40, TEX, DOL, VIC, and HOB each occurred on 60% of the trials while LIK, CEN, BAL, and SUR each occurred on 40% of the trials. For the other half of the *Ss* TEX, DOL, VIC, and HOB each occurred on 40% and LIK, CEN, BAL, and SUR each occurred on 60% of the trials. Thus, relative percentages of occurrence of TEX, DOL, VIC, and HOB and of LIK, CEN, BAL, and SUR were counterbalanced. For the other RP, occurrences of members of these two sets of alternative responses were counterbalanced in the same way.

There were two constraints on the randomness of occurrences of alternative response members of each unit. One constraint was that the particular RP would be realized for each unit within each of the two groups of 10 permutations of orders of occurrence of stimulus members. The other constraint was that the patterns with which these RP were realized were the same through each of the six repetitions of each of the two groups of 10 permutations.

(*b*) *One-unit.* For each RP, there were eight one-unit lists, each of which was one of the units of the four-unit list with one or the other response member occurring more frequently. For the 50–50 RP, the order of the alternative response members was interchanged. Each list was administered to a different *S*.

(*c*) *One-unit spaced.* Length and variability of time intervals between successive presentations of the unit were the only differences between one-unit and one-unit spaced conditions. These intervals were obtained for each of the eight lists by inserting blank cards for the three PA units of the four-unit list which were not being used.

(*d*) *One-unit random.* Instead of the blank cards of the one-unit spaced condition there were pairs of syllables which had been formed by drawing 30 successive pairs of syllables at random from decks of 60 syllables. Each deck consisted of the nine syllables which were the members of the three units not being used as the constant unit. There were 10 instances of each of three syllables which were stimulus members of these three units and five instances of each of the six syllables which were response members. Within the 30 pairs, each of the nine syllables was paired with each of the other eight syllables, and each was, on the average, a stimulus member half of the time and a response member half of the time.

Apparatus and procedures.—Each unit was presented at a 2:2 sec. rate by means of a Hunter Card Master. Each exposure of the blank cards of the one-unit spaced lists was for 4 sec. The interunit interval of the four-unit, and one-unit random conditions was 2 sec. The interunit interval and the intervals between units and blank cards or between blank cards of the one-unit spaced condition were each 2 sec., but counting the time required for occurrences of blank cards, time intervals between successive presentations of the single unit ranged from 2 to 38 sec. about a mean of 20 sec.

Each of the units of the four-unit list and the single unit of the one-unit, one-unit spaced and one-unit random conditions was presented 120 times. Under the one-unit spaced condition, the 120 presentations of a unit alternated with 360 presentations of blank cards. Under the one-unit random condition, the 120 presentations of a unit alternated with 360 presentations of the randomly-formed pairs.

The same essentially standard instructions for paired-associates learning, with anticipations spelled, were given to *Ss* in each of the 24 combinations of conditions. They were not told that each stimulus member might have alternative response members.

RESULTS

For RP from 60–40 to 100–0, the response measure was percentage of anticipations of the more frequent and less frequent response member of each unit in each of six successive blocks of 20 trials each. During the last 60 trials under all four list conditions and essentially for all trials with the

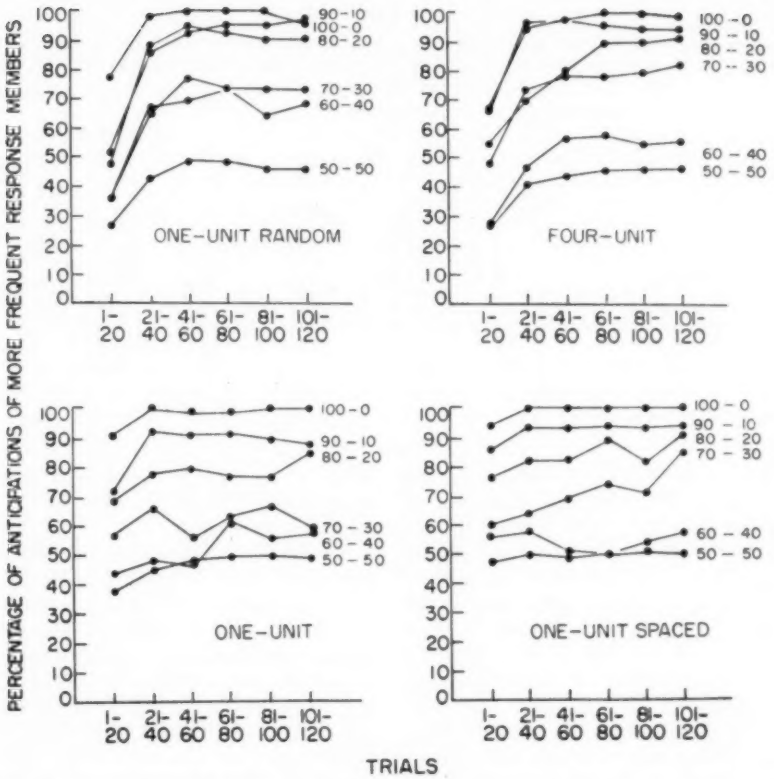

FIG. 1. Mean percentages of anticipations of the more frequent response members for successive 20-trial blocks for each of the 24 combinations of RP and L.

one-unit spaced and one-unit random lists percentages of anticipations of the more frequent and less frequent response members for each percentages summed almost to or to 100%. For this reason only percentages of anticipations of the more frequent members are presented. For the 50–50 RP, the measure for each unit was the average of the percentages of anticipations of both response members of the unit in each successive block of 20 trials. For each *S* administered the four-unit list, percentages of, anticipations of response members for each of the four units were averaged to obtain a single percentage for each block.

Over-all differences among combinations of conditions.—Figure 1 shows means of percentages of anticipations of more frequent response members for successive blocks for each of the 24 combinations of RP and Lists (L). Differences among these trends for all six and during the last three blocks of trials (T) were assessed by the analyses of variance summarized in Table 1.[3]

[3] With the 100–0 relative percentages, *Ss* responded almost perfectly from Trials 21–40 on. In order to save both some *Ss'* and *Es'*

TABLE 1

ANALYSES OF VARIANCE OF PERCENTAGES OF ANTICIPATIONS OF MORE FREQUENT RESPONSE MEMBERS FOR ALL SIX BLOCKS OF TRIALS AND DURING THE LAST THREE BLOCKS OF TRIALS

Source	All Six Blocks			Last Three Blocks		
	df	*MS*	*F*	*df*	*MS*	*F*
RP	5	77,230.60	178.51**	5	35,422.54	171.84**
L	3	1,030.33	2.38	3	990.24	4.32**
RP × L	15	723.33	1.67	15	483.86	2.11*
Error (b)	168	432.63		168	229.41	
T	5	12,458.15	146.85**	2	105.03	2.53
T × RP	25	161.90	1.91**	10	51.41	1.24
T × L	15	1,458.98	17.17**	6	77.38	1.87
T × RP × L	75	122.08	1.44**	30	54.37	1.31
Error (w)	810	84.95		306	41.47	

* $P < .05$.
** $P < .01$.

For all six blocks of trials, *F*s significant at or beyond the .01 level were obtained for RP, T, T × L, T × RP, and T × RP × L. The significant T × RP × L interaction reflects three aspects of the relationship among the curves of Fig. 1. One aspect is that which produced the T × L interaction. For Trials 1–20, percentages of anticipations for four-unit and one-unit random lists were markedly below those for one-unit and one-unit spaced lists. By Trials 21–40 and thereafter, the percentages for the former two lists were slightly above those for the latter two lists. When averaged across all 120 trials, the early lower and subsequent higher values of the percentages for four-unit and one-unit random lists yielded percentages across all 120 trials which were about the same as those across all 120 trials for the one-unit and one-unit spaced lists. Thus there was a T × L interaction, but *F*s for L and RP × L were not significant.

The second aspect is that which produced the T × RP interaction. The increments in percentages of anticipations from Trials 1–20 to subsequent trials were successively greater in negatively accelerated fashion for RP from 50–50 to 80–20 or 90–10, but the increment for 100–0 was only about equal to that for 60–40; thus, there was an inverted-U relationship between these increments and RP. The third aspect is that, unlike the curves for RP from 60–40 to 90–10, the curves for 50–50 for four-unit and one-unit random lists tended to be slightly below those for one-unit and one-unit spaced lists even during the last 60 trials. For 100–0, all four curves were about equal from Trials 41–60 onward.

The *F* for RP reflects a direct relationship between percentages of anticipations and actual percentages of occurrence which held not only across but also within combinations of L and T.

For the last three blocks of trials, neither the *F* for T nor *F*s for inter-

time, the last 40 trials were eliminated for 15 *S*s who were under 100–0 RP with four-units, one-unit spaced, and one-unit lists. For these three combinations, the *N*s for those blocks of trials were reduced by 4, 6, and 5, respectively. Missing values for the last two blocks were estimated from cell means; the *df*'s of "within *S*s" error variances were reduced by 30.

TABLE 2

MEANS AND *SD*s OF PERCENTAGES OF ANTICIPATIONS OF MORE FREQUENT RESPONSE MEMBERS DURING TRIALS 61–120

Lists (L)	Relative Percentages of Occurrence (RP)											
	50–50		60–40		70–30		80–20		90–10		100–0	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Four-unit	46.5	2.1	56.4	11.6	80.0	10.0	91.0	5.6	95.3	2.5	99.5	0.4
One-unit random	47.0	4.4	69.6	19.5	74.5	15.5	91.6	5.5	96.9	3.6	99.2	2.3
One-unit spaced	49.8	0.4	53.8	9.4	76.5	7.8	87.3	7.9	94.0	4.4	100.0	0.0
One-unit	49.8	0.3	58.7	6.6	63.0	13.9	80.0	7.6	90.2	9.7	99.8	0.2

actions involving T were significant. The RP × L interaction indicates that rank-orders of the four list conditions, differences among those conditions, or both differed among RP. For 50–50 and 100–0, there were only slight differences among the four list conditions. For the remaining RP, with some inversions, the rank-order of decreasing percentages of anticipations was one-unit random, four-unit, one-unit spaced, and one-unit lists. The differences among list conditions for these RP were sufficient to produce the significant F for L. The significant F for RP reflects a direct relationship between percentages of anticipations and actual percentages both across and within list conditions.

Since there were no further changes in mean percentages of anticipations of the more frequent response members during the last 60 trials, means of percentages of anticipations for Trials 61–120 were computed as single, more reliable estimates of asymptotic percentages (Table 2). Rates of approach (θ') to these presumed asymptotes were estimated by the equation derived by Voss (1958).[4] Table 3 presents the values of θ' obtained for all but the one-unit spaced, 60–40 combination. (Because of a decrease between earlier and terminal percentages, the θ' for this combination could not be computed.) For each RP, the θ's for the three one-unit conditions were, with two exceptions, larger than the corresponding θ' for the four-unit list. Within each list condition, with some inversions, values of θ' first decreased from those at 50–50 and then increased to highest values at 100–0.

Differences among specific combinations of conditions.—Mean percentages of anticipations during Trials 61–120 were also used for comparisons among specific combinations of conditions.

TABLE 3

VALUES OF θ' FOR COMBINATIONS OF RP AND L

Lists (L)	RP					
	50–50	60–40	70–30	80–20	90–10	100–0
Four-unit	.085	.075	.102	.067	.194	.136
One-unit random	.113	.100	.087	.109	.085	.234
One-unit spaced	.678	—	.106	.204	.470	.800
One-unit	.131	.079	.296	.305	.305	.544

[4] The equation used was

$$\theta' = p(\infty)/np(\infty) - \sum_{1}^{n} \overline{p(n)}$$

where θ' is the estimate of learning rate, $p(\infty)$ is the estimated asymptote of percentages of anticipations of the more frequent response member, n is the number of trials, and $\sum_{1}^{n} \overline{p(n)}$ is the total number of occurrences of the more frequent response in n trials.

TABLE 4

DIFFERENCES (d) BETWEEN PERCENTAGES OF ANTICIPATIONS OF MORE FREQUENT RESPONSE MEMBERS AND THEIR ACTUAL PERCENTAGES OF OCCURRENCE DURING TRIALS 61–120 AND t'S FOR THOSE DIFFERENCES

Lists (L)	Relative Percentages of Occurrence (RP)									
	50–50		60–40		70–30		80–20		90–10	
	d	t	d	t	d	t	d	t	d	t
Four-unit	−3.5	4.38**	−3.6	0.82	+10.0	2.64*	+11.0	5.16**	+5.3	5.64**
One-unit random	−3.0	1.86	+9.6	1.30	+4.5	0.77	+11.6	4.11**	+6.9	5.06**
One-unit spaced	−0.2	0.48	−6.2	1.76	+6.5	2.22	+7.3	2.44*	+4.0	2.45*
One-unit	−0.2	1.39	−1.3	0.53	−7.0	1.33	0.0	0.0	+0.2	0.07

* $P < .05$, $df = 7$.
** $P < .01$, $df = 7$.

Mean percentages for 100–0 were excluded from these comparisons to reduce heterogeneity of variances, and to eliminate the problem of some estimated percentages over the last 40 trials. Also, the 100–0 RP differed from the others: while the 100–0 could only be below or at 100%, the others could be below, at, or above the percentages of occurrence of the more frequent response member.

A Duncan multiple range test (Federer, 1955, pp. 26–29) indicated that for RP of 50–50, 80–20, and 90–10, there were no differences among the four lists conditions. For 60–40, the percentage for the one-unit random list was significantly larger than that for the one-unit spaced list. For 70–30, the percentage for the one-unit list was significantly smaller than the percentages for one-unit spaced and four-unit lists. In general, percentages of anticipations for a particular list condition within each RP did not differ significantly from percentages of anticipations for that condition within RP 10% removed.

Percentages of anticipations compared with percentages of response-member occurrence.—For each RP except 100–0, Table 4 presents t's for differences between mean percentages of anticipations of more frequent response members and actual percentages of occurrence of those members. The error terms of these t's were computed from the *SD* for each particular combination. Since Wilkinson's (1951) table indicates that eight of 20 t's significant at the .05 level would occur less than once in a 1000 times, the significant t's as a whole cannot readily be attributed to chance factors. For the four-unit list with RP of 50–50, the percentage of anticipations was significantly lower than 50%. All other significant t's are for percentages of anticipations of more frequent response members which exceeded their actual percentages of occurrence. For the four-unit list, differences for 70–30, 80–20, and 90–10 were significant at from .05 to less than .01. For both one-unit spaced and one-unit random lists, the differences for 80–20 and 90–10 were significant at from .05 to less than .01. No difference for the one-unit list approached significance.

Differences between sets of more frequent response members and among units of the four-unit list.—Within the four-unit list, TEX, DOL, VIC, and HOB occurred both more frequently and less frequently than their respec-

tive alternative response members, LIK, CEN, BAL, and SUR. An analysis of variance revealed no difference between the two sets of more frequent alternative response members.

During the first three blocks of trials, analyses of variance indicated that there were differences among units for both more frequent and less frequent occurrence of TEX, DOL, VIC, and HOB. However, during the last three blocks, the units were equivalent.

Discussion

Three features of the results of this experiment are of particular importance. One is the effects of RP and L on rates of approach to and apparent asymptotes of percentages of anticipations of more frequent response members. The second is implications of the present results for the generality of theory and findings for two-choice situations with light stimuli and a single unit. The third is the possible use of divergent paired-associates units to establish hierarchies of stimulus-response associations.

Percentages of anticipations of the more frequent response member.—Over the last three blocks of trials, none of the *F*s involving trials was significant. Thus, for each of the 24 combinations of RP and L, asymptotic percentages of anticipations were apparently reached on or before Trials 61–80. Only with certain combinations of RP and L did asymptotic percentages of anticipations of the more frequent response equal their actual percentages of occurrence.

For 50–50, with both one-unit and one-unit spaced lists, the asymptotic percentages of anticipations were almost exactly 50%. But the estimated asymptote of 46.5% with the four-unit list was significantly below 50%, and the decrease from 50% with the one-unit random list, though not statistically significant, was almost as great. The remaining responses were sometimes intralist intrusions which indicates that these apparent decrements were largely due to interference from responses to the stimuli of other units of the four-unit and one-unit random lists.

For 60–40, the smallest difference between percentages of anticipations and actual percentages was with the one-unit list. Even though the absolute values of the other three differences were up to 9.6%, none was statistically significant. With the one-unit list, only with RP of 70–30, was there a sizeable difference between percentages of anticipations and actual percentages. But this difference, in the direction of fewer than 70% anticipations of the more frequent response member, was not statistically significant. The other three differences represented anticipations of the more frequent response member exceeding 70%, but only that for the four-unit list was statistically significant.

For 80–20 and 90–10, with the one-unit list the percentages of anticipations of the more frequent response were almost exactly equal to 80% and 90%. But all of the percentages of anticipations with the other three lists were significantly larger than 80% and 90%. Thus, for all but the one-unit list, there is apparently some point between 60–40 and 80–20 at which percentages of anticipations of the more frequent response member begin to exceed their actual percentages of occurrence. Occurrences of the alternative responses of each unit were semirandom rather than random, a circumstance which should facilitate matching actual percentages. Accordingly, those higher percentages of anticipations than of actual percentages which were obtained may represent minimum rather than maximum differences.

For almost all RP, percentages of anticipations with the four-unit and one-unit random lists seemed more similar to each other than to those with the one-unit spaced list which differed somewhat from the one-unit list. The former difference suggests that, whether the pairings of the stimuli between a particular unit were constant or random, the stimulation and activities associated with occurrences of those units had some differential effect on anticipations rela-

tive to repetitions of a particular single unit separated by the same longer, variable intervals filled with occurrences of blank cards. For RP of 80–20 and above, the longer, variable intervals seem to occasion percentages of anticipations greater than the matching percentages obtained with the one-unit list.

For 70–30, 80–20, and 90–10 with four-unit, one-unit random, and one-unit spaced lists, the results of the present experiment agree closely with Voss et al. (1959) findings of percentages of anticipations of more frequent response members larger than their actual percentages of occurrence. For these three list conditions, which are more similar to Voss et al. lists than is the one-unit list, there is only one pronounced deviation from their results: for 60–40, percentages of anticipations with four-unit and one-unit spaced lists were below rather than above 60%. This disagreement may reflect chance factors, and hence is only apparent, or it may represent replicatable effects of one or more differences in stimuli, apparatus, and procedures.

Generality of equations.—Further use of equations such as Estes' (1959b, p. 411) for predictions of acquisition with divergent PA units, without some modification, apparently depends on the particular combinations of RP, number of units, and conditions of presentation of one-unit lists. With the one-unit list, asymptotic percentages for all RP and actual percentages did not differ significantly and only with 70–30 was the value of the difference of any magnitude. Thus, the equation seems applicable without modification. However, this relatively close fit of anticipations and actual occurrences has one perplexing feature: a closer fit would probably have been predicted with the longer, variable interunit intervals of the one-unit spaced list than with the short, interunit interval of the one-unit list (1959b, p. 421), while the results for the two lists were just the opposite. Conceivably the two conditions would not have had different effects had *Ss* merely waited during the longer, variable intervals rather than looking at blank cards.

With the one-unit spaced and one-unit random lists for RP of 80–20, 90–10, and perhaps of 70–30, and with the four-unit list for RP of 70–30, 80–20, and 90–10, asymptotic percentages of anticipations of the more frequent member exceeded actual occurrences. Furthermore, with the four-unit list and perhaps the one-unit random list for RP of 50–50, asymptotic percentages were apparently less than 50%. Thus, any satisfactory equation must provide for predictions of asymptotic percentages of more frequent response members which, for specific combinations of RP and L, are less than, equal to, or greater than actual percentages of occurrence of those members. For this objective, even more recent extensions based on "component and pattern models" (Estes, 1959a) seem inadequate. The rate parameter also seemingly varied with RP and L in a manner roughly consistent with Voss et al. findings. Therefore, either values for this parameter must be determined empirically for each specific combination of RP and L, or an equation expressing θ' as a function of RP and L must be developed.

As predicted by equations and findings for two-choice tasks involving simpler stimuli and a single unit, the curves for most of the combinations of RP and L were negatively accelerated. The exceptions were some for RP of 50–50, 60–40, and 70–30 with one-unit and one-unit spaced lists whose initial negative acceleration was completed so early in the first 20 trials, that their means for those trials were essentially at asymptotic percentages.

Hierarchies of stimulus-response associations.—The divergent PA unit may prove useful for experimentally establishing hierarchies of associations of two or more responses to what is ordinarily conceived as the same or a single stimulus. (Analytically, such stimuli may best be treated as constituted of n elements or as appearing in one or more of n patterns each of which is associated with one response in all-or-none fashion.) At present, hierarchies of stimulus-response associations or of responses are usually inferred from hierarchies of responses to a stimulus

for large groups of *Ss*. On the assumption of the existence of hierarchies of responses to a single stimulus, a procedure for establishing such hierarchies on the part of individual *Ss* would have two advantages: (*a*) inferences from hierarchies for groups to those for individual *Ss* could be replaced by more direct and presumably more precise specification of hierarchies for individual *Ss*, and (*b*) with different hierarchies of responses to the same stimulus, confounding of different stimuli with different hierarchies could be eliminated.[5]

Following a suggestion by Hull (1930, pp. 244–245), Peterson (1956) tested a less direct procedure for establishing hierarchies of responses to stimuli. The stimulus members of Peterson's paired-associates were compounds of two nonsense syllables. One response was conditioned to one syllable of the compounds and two responses were conditioned to the other syllable of the compounds. After the learning criterion has been attained, the syllables of the compounds were presented separately to permit determination of frequencies of alternative responses to the syllable to which those responses had been conditioned. A disadvantage of this procedure is that acquisition of potential hierarchies cannot be observed except by interpolated test trials. Also, although Peterson was able to predict observed responses from his formula, the procedure involves the possible complication of adjustments for training with a stimulus in a compound and testing with the stimulus alone (Hull, 1943, p. 220).

Osgood and Anderson's (1957) task could be viewed as consisting of 10 paired-associates units each of which had three response members which occurred with percentages of .60, .30, and .10. However, since each name was both a stimulus member and a response member, because of "backward associations," it is difficult to specify the potential number of responses in a hierarchy. Thus, the name "Josephine" of Osgood and Anderson's illustration might be regarded as having not three but six different responses.

[5] Estes (1960, p. 212) has recently presented analyses of responses after two or four trials with divergent PA units the pattern of which was interpreted as evidence against an assumption of an hierarchy of responses to a single stimulus for individual *Ss*. Until data for test-trial performance following more than one or two occurrences of the alternative response members are available, both the generality of those findings and his interpretation seem questionable. Examination of the data for the first 20 trials for individual *Ss* of the four-unit 50–50 combination of the present experiment disclosed that, in 18 of 32 instances, one or the other response member was anticipated on 75% of the occurrences of a particular unit. Thus, Estes may simply have demonstrated that one or the other of the alternative stimulus-response relationships of a divergent unit, presumably because of pre-experimental experiences, has considerably and consistently greater initial strength.

Summary

Acquisition with divergent paired-associate (PA) units was investigated as a function of six relative percentages of occurrences of alternative response members of a unit (RP) which varied from 50–50 to 100–0 in 10% steps, and of number of units, four or one. In order to control for effects of time intervals between presentations of the same unit in the four-unit list and for events within those intervals, there were three conditions of presentation of single units: (*a*) a constant, 2-sec. interval between each presentation; (*b*) variable, longer intervals between each presentation, and (*c*) the same variable, longer intervals filled with presentations of randomly-paired stimuli. Eight undergraduate *Ss* were assigned to each of the 24 combinations of RP and L wherein they had 120 trials with each unit of the four-unit list or with the single unit of the three one-unit conditions.

Across all six blocks of 20 trials each, and during the last three blocks, there was a direct relationship between percentages of anticipations of the more frequent response members and their actual percentages of occurrence. During the last three blocks, decreasing percentages of anticipations were obtained with one-unit random, four-unit, one-unit spaced, and one-unit lists. Differences in rate of acquisition were reflected in significant interactions of the six blocks of trials with RP, with L, and with RP and L.

The RP × L interaction for the last three blocks of trials could be interpreted as reflecting differences among lists with respect to the degree to which percentages of anticipations of the more frequent response members approximated their actual percentages of occurrence. With the one-unit list, percentages of anticipations were essentially the same as actual percentages. With four-unit, one-unit random, and one-unit spaced lists for 80–20 and 90–10, and probably for 70–30, percentages of anticipations exceeded actual percentages.

The significance of these results for findings and theory for apparently homologous situations involving simpler stimuli was discussed. Also noted was the possible use of divergent PA units for establishing response hierarchies experimentally.

References

ARCHER, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

ESTES, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. (a)

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. (b)

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67,** 207–223.

FEDERER, W. T. *Experimental design: Theory and application.* New York: Macmillan, 1955.

HULL, C. L. Simple trial and error learning: A study in psychological theory. *Psychol. Rev.*, 1930, **37,** 241–256.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

OSGOOD, C. E., & ANDERSON, L. Certain relations among experienced contingencies, associative structure, and contingencies in encoded messages. *Amer. J. Psychol.*, 1957, **20,** 411–420.

PETERSON, L. R. Prediction of response in verbal habit hierarchies. *J. exp. Psychol.*, 1956, **51,** 249–252.

VOSS, J. F. Supplementary report to "language behavior and subjective probability." *Nat. Inst. Ment. Hlth. Rep.*, 1958, No. M-1476.

VOSS, J. F., THOMPSON, C. P., & KEEGAN, J. H. Acquisition of probabilistic paired associates as a function of S-R_1, S-R_2 probability. *J. exp. Psychol.*, 1959, **58,** 390–399.

WILKINSON, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48,** 156–158.

(Received April 25, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 35–42

SEPARATING PERCEPTUAL AND LINGUISTIC EFFECTS OF CONTEXT SHIFTS UPON ABSOLUTE JUDGMENTS[1]

DAVID L. KRANTZ AND DONALD T. CAMPBELL

Northwestern University

In accordance with the usual strategy of parsimony in science, the effects of extreme contexts upon absolute judgment were interpreted initially as illustrating a *single* process of relativity of judgment or adaptation level. The most economical interpretation of the data to be presented here necessitates the postulation of *two* such processes, both producing contrast effects in the same direction. One of these processes, transient and reversible, is found in both of two judgmental languages, and is designated *perceptual*. The other, irreversible and specific to an unfamiliar, experimenter-defined response language, is designated *semantic*.

Two papers published in 1958 independently suggest dividing the judgmental phenomena currently grouped under Helson's concept of adaptation level (e.g., Helson, 1947, 1948, 1959) into semantic or linguistic effects on the one hand, and perceptual or end-organ effects on the other. One of these papers (Campbell, Lewis, & Hunt, 1958) argued that the "situationally relative," "novel," "arbitrary" and "restricted" response language ordinarily used in the method of single stimuli could well have produced the contrast effects resulting from extreme anchors or extreme contexts without generating distortions in identity judgments (had they been called for) or any perceptual illusion. That study then examined the effect of a shift in context of stimulation upon the judgment of a common stimulus, employing a judgmental language which was "absolute, extensive, and extra-experimentally anchored." Finding the usual contrast or adaptation level effects, the paper concluded that such phenomena could not be explained away as semantic artifacts. However, it did not raise the possibility that both semantic and perceptual effects are present in the typical judgmental assignment.

The other paper suggesting a duality of processes was that of Stevens (1958) in which he argued that different phenomena were involved in sensory adaptation (in which end-organ receptor-cell adaptation takes place) on the one hand, and in judgment (where redefinition of categories is involved) on the other, although previous approaches treat the two as if the same process were involved. Here again, two processes were postulated with no explicit assumption that both could be present in the same judgmental task. The outcomes of the present study seem to require such an assumption.

The setting of the present study is the judgment of the lengths of white lines projected on a screen. A common "tracer" stimulus recurs in all phases of the experiment, in one phase interspersed among longer lines, in another phase occurring among shorter lines. One set of *S*s judges in "inches," the other uses an *E*-defined

[1] This study is based upon an MA thesis done by the first author under the second author's direction. Certain analyses beyond those reported in the thesis (Krantz, 1959) have been financed by ONR Project 7onr-450(11), W. A. Hunt, Principal Investigator. The authors are also indebted to Allen Parducci for editorial suggestions.

response language for which the value "100" is "average." The effects of the shift in context upon judgments of the common tracer stimulus are examined.

METHOD

Slides were prepared which projected white lines of 13 different horizontal lengths, for a .2-sec. duration, with a 2-sec. interstimulus interval. The slides were made by using black tape on clear glass $3\frac{1}{4} \times 4$ in. slides. Projected on the screen, the shortest line (No. 1) was approximately 6 in., the longest 36 in., the middle "common" or "tracer" stimulus (No. 7) 20 in. (the exact values of the stimuli are presented in Table 1) and the vertical widths $3\frac{3}{4}$ in. The slides were organized in groups of seven (here called "trials") within each of which the common tracer Stimulus 7 occurred once, the order of stimuli within a trial being random. For the Low Phase, the stimuli in each trial ranged from Stimulus 1 to Stimulus 7. For the High Phase, the range was from Stimulus 7 to Stimulus 13. Trials in the Transition Phase also spanned seven steps, graduating between the Low and High Phases. The three transition trials ranged between Stimuli 3–9, 4–10, 5–11 for Low to High and vice versa for a High to Low transition. Both a High-Low and a Low-High group were run for each combination of the other three experimental variables.

In the second experimental treatment variable, the "Inches" condition was presented to *Ss* with these instructions, "You are to judge these lines in terms of how long, in inches, they appear to be." The "Rating" condition was introduced this way:

> You are to judge these lines in terms of a rating scale which you will construct during the course of the experiment. When you see a line which appears "average in length" to you, judge that line with the number 100. If a line appears less than average, assign the number 95. If another line appears even smaller than the average line and smaller than the 95 one, use the number 90 to designate "less, less than average."
>
> In the same way, if a line is above average in length, assign the number 105. If another line is larger than the 100 average line and larger than the 105 line, assign the number 110. In other words, you can build your rating scale above and below the average rating of 100. You are not restricted to the numbers I have given you. The numbers I have given you on the reference sheet are only examples. In-between numbers, like 93 or 106 are appropriate. You can use numbers less than 90 or greater than 110. In other words, you are constructing a scale which shows the differing lengths of lines presented. Any system of numbers, such as in-between numbers or numbers above 110 or below 90 are usable. The only stipulations are that you use 100 as average and that if you assign a number to a particular length of line, the next time that line is presented you assign it the same number. Let me repeat, always use the same number for the same length of line; do not change your judgment of the same line.
>
> Let me review the categories or numbers on your reference sheet. The number 100 stands for the average line; 95 for less than average; 90 for less, less than average; 105 for more than average; 110 for more, more than average.

In choosing this particular form of rating, there was an effort to eliminate the effects of "vocabulary exhaustion" (Campbell, Lewis, & Hunt, 1958) such as occurs when all of a restricted number of ratings have been employed prior to a shift of context, and to exclude a deliberately relative or comparative usage by *Ss* which would allow them consciously to redefine values in the course of the experiment.

The third experimental variable employed was Preshift Training, i.e., the length of the initial phase, presumably manipulating the persistence of the adaptation level established therein. "Phase 1" indicates the initial set of trials given, irrespective of the context employed. This period was followed by the transition series ("Phase 2"). A final series of trials after shift of context is designated "Phase 3." In the Short condition, five Phase 1 trials were used; in the Long condition, 15.

The fourth variable was the degree of illumination, which was presumed to affect stimulus discriminability. Under the Bright condition, the projected lines registered .73 apparent ft-c. Under the Dim condition a neutral density filter reduced the brightness to .32 apparent ft-c. The background brightness of the screen was .26 apparent ft-c for both illumination conditions.

These four dichotomies generate 16 experimental treatments. For each of these, 20 *Ss* were obtained, a total of 320 *Ss*. The *Ss* were drawn from introductory classes in psychology. They were tested in groups of four to six, seated in two rows, 13 and 16 ft. from the screen. Experimental sessions were randomly assigned to treatments, with the restriction that all 16 treatments be evenly represented during all periods of the three month course. In two successive quarters the experimental design was replicated with 10 *Ss* in each cell.[2]

The *Ss* recorded their own judgments in a

[2] One session of 6 *Ss* (in treatment High-Low, Rating, Short training, Dim illumination) was discarded because of extremely aberrant results. This session was subsequently rerun twice with different *Ss* to confirm the aberrance, attributable to intra-experimental history (Campbell, 1957). It is felt that this substitution of the rerun group data augments rather than reduces the replicability of the study's findings.

TABLE 1

MEAN JUDGMENT OF EACH STIMULUS DURING FIRST FIVE TRIALS OF PHASE 1

Stimulus Number	Low Context			Stimulus Number	High Context		
	Length (In.)	Mean Judgment: Inches	Mean Judgment: Ratings		Length (In.)	Mean Judgment: Inches	Mean Judgment: Ratings
1	6.1	3.74	79.90	7	20.2	12.74	90.98
2	9.0	5.37	88.99	8	23.0	14.38	94.00
3	11.0	6.75	92.48	9	25.0	16.95	97.78
4	12.4	7.80	96.09	10	28.1	20.90	101.74
5	15.8	10.77	102.55	11	31.3	23.32	103.72
6	18.7	13.32	106.29	12	33.1	27.16	108.21
7	20.2	17.21	111.77	13	35.8	31.59	112.63
Mean		9.28	96.87	Mean		21.01	101.29

Note.—Preshift training and discriminability conditions pooled, $N = 80$ *Ss* per mean.

booklet of slips with each page allowing 10 judgments. Twelve *Ss* were discarded for getting out of step with the stimulus presentation and for incomplete records. Excess *Ss* over the 20 required for each treatment were discarded by the use of random numbers.

RESULTS AND DISCUSSION

While the statistical analyses to be presented have been done separately for each judgment language, in Fig. 1 and 2, there is an effort to show both sets of results on the same graph. This necessitates computing a conversion factor between the two judgment languages, which in turn necessitates examining data on judgments of stimuli other than Stimulus 7, data not otherwise used. Because of their relevance to the attempted translation between the two languages, and to the more general problem of the relationship of stimulus and judgmental distances, these data are presented in Table 1.

The data do not lead to any univocal conversion ratio. If ratings be plotted against stimuli, the plots are satisfactorily linear, but of differing slope for the two contexts, the High context being the flatter. If judged inches be plotted against stimuli, the slopes depart slightly from linearity in the unexpected direction of positive acceleration for both High and Low contexts, with the High context being the steeper in this case. When rating means are plotted against judged inches, the Low context shows a slight bowing, and a steeper slope than the High context. As an approximate average of the slope of the two contexts, a ratio of 1 in. = 1.50 rating points has been chosen for the purpose of graphing Fig. 1 and 2. The graphs have been centered on the mean for the two contexts of the column means of Fig. 1, i.e. 15.64 in. and 99.08 rating points. Because of the differing slopes for the two contexts, this represents a more arbitrary decision than had been hoped for.

In Fig. 1 and 2, all data represent judgments of the same tracer stimulus, No. 7, and all connected points represent judgments by the same 40 *Ss*. From the Phase 1 differences between the corresponding High-Low and Low-High groups, one can see that the expected contrast effects emerge. The striking shifts in each line between Phase 1 and Phase 3 are likewise uniformly in the expected direction. But perhaps most dramatic is the difference in overall pattern between the Inches and Ratings conditions, a difference which provides the central problem of interpretation.

A summary of the results of six independent analyses of variance, shown in Table 2, provides statistical

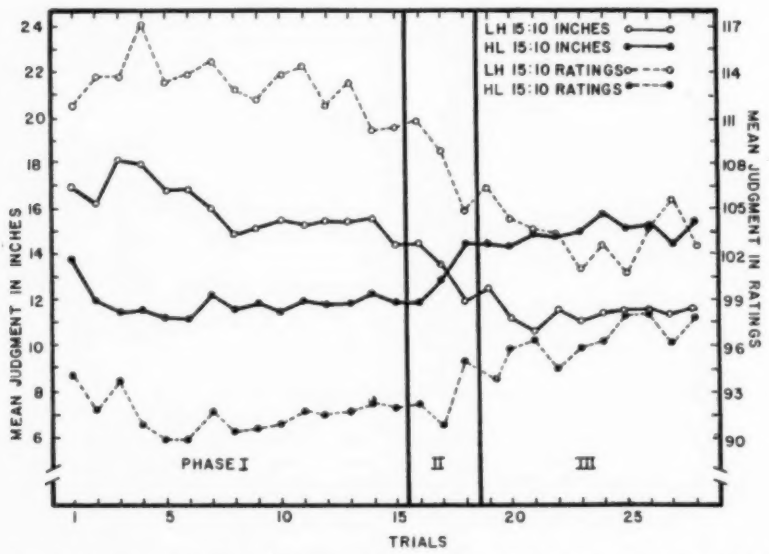

FIG. 1. Mean trial-by-trial judgment of tracer Stimulus 7 for Long training condition. (Discriminability conditions pooled, $N = 40$ for each point.)

confirmation of the overall pattern for Inches and Ratings conditions. In the Phase 1 analyses, the mean of each S's judgment of Stimulus 7 during the last four trials of Phase 1 has been used, since this set of responses is comparable across Preshift Training conditions. For both response languages, the main effect of context is highly significant and is the only significant effect. As graphically illustrated in Fig. 1 and 2, the effect is greater for the Ratings. Discussion of outcomes such as these frequently imply that the difference between the F ratios is itself significant. Rarely, however, is this higher-order significance tested. Bradley and Schumann (1957) have recently provided an F test among Fs appropriate for parallel experiments such as the present one. By their criteria ($w' = 21.32$, $a' = 33.47$, $b = 76$, $P < .05$, minimum $w'_{.05} = 1.85$),[3] the effect is significantly greater for the Ratings condition. This is one of the major findings of this study.

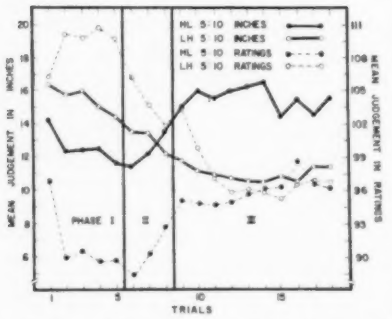

FIG. 2. Mean trial-by-trial judgment of tracer Stimulus 7 for Short training condition. (Discriminability conditions pooled, $N = 40$ for each point.)

[3] The tables supplied by Bradley and Schumann provide w' values only at 5% level of significance. These minimum w' values will be presented in parenthesis following the obtained values. Note that the obtained values greatly exceed these minimums.

The Phase 3 analyses are based upon the means of each *S*'s judgments of Stimulus 7 in the 10 Phase 3 trials. Here the effect of context is highly significant for the Inches condition. (That the *F* of Phase 3 Inches seems more significant than that of Inches in Phase 1 is due primarily to reduced variability which is presumably attributable to the greater amount of practice and to scores based upon 10 judgments rather than 4.) As seen in Fig. 1 and 2, the effect of context on Ratings is weak and in the opposite direction from that on Inches. This main effect ($P < .05$) is graphically clear only for the Long Preshift Training. The difference in Phase 3 context effects *F*s between Ratings and Inches is highly significant by the Bradley and Schumann (1957) test ($w' = 7.94$, $a' = 4.23$, $b = 76$, $P < .05$, minimum $w'_{.05} = 3.18$).

Note the apparent inconsistency of the Phase 1 and Phase 3 findings. For Phase 1 both response languages show context effects in the same direction with Ratings being larger. For Phase 3 the context effects are in different directions, with Inches being greater. In the graphic presentation of trial-by-trial results of Fig. 1 and 2, the Inches judgments show complete reversal with the reversal of contexts, the lines crossing in the transition phase. In contrast, the Ratings never cross over, the direction of difference in Phase 3 remaining the same as in Phase 1.

Minor Findings.—In the Phase 3 analysis, the effect of Preshift Training is significant at the marginal 5% level for the rating groups. Those groups with shorter Phase 1's give lower ratings in Phase 3. This main effect is almost wholly contributed by the Low-High groups with the Short Phase 1.

The findings of the Phase 1-Phase 3 Shift analysis are not to be confused with a test of the significance of the Phase 1-Phase 3 shift, the uniformly high significance of which is reported below. Rather, the Phase 1-Phase 3 shift analysis shown in Table 2, examines the effects of treatments other than Phase upon the amount of shift. The score employed for each *S* is the difference in mean judgment between the last 4 trials of Phase 1 and the 10 trials of Phase 3, the direction of subtraction being reversed between High-Low and Low-High groups so that all changes in the expected contrast direction are positive. The analysis was designed primarily to demonstrate the expected effects of Discriminability and Preshift Training on stability of the adaptation level, or upon the differential occur-

TABLE 2

F VALUES FROM ANALYSES OF VARIANCE

Source	*df*	Phase 1 Judgments		Phase 3 Judgments		Phase 1-Phase 3 Shift	
		Inches	Ratings	Inches	Ratings	Inches	Ratings
Order (context) (O)	1	11.95**	254.79***	40.99***	5.16*	2.42	18.48***
Preshift training (T)	1	.79	2.27	.29	5.12*	1.99	4.16*
Discriminability (D)	1	1.40	1.34	.47	3.47	1.22	.22
O × T	1	.05	.02	.52	2.66	.54	1.85
O × D	1	.55	1.02	.01	.43	1.14	1.08
T × D	1	.54	.34	.23	.37	.97	.57
O × T × D	1	.00	.25	.20	.62	.02	.01
Within groups (*MS*)[a]	152	(23.20)	(60.35)	(21.52)	(84.27)	(8.72)	(42.98)
Inhomogeneity of variance	1/20	1.60	1.85	2.42	1.24	2.79	4.87*

[a] Mean squares shown in parentheses, rather than *F*s.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

rence of assimilation and contrast illusions among *Ss*. Of the 320 *Ss*, only 20 had negative scores and these were individually insignificantly different from no shift at all. Thus, this setting produced no persons who were significant assimilators, unlike two other settings in which these effects have been examined (Campbell, Hunt, & Lewis, 1957, 1958; Campbell et al., 1958).

The significant main effect for Order for shift scores is conceptually an interaction with the counterbalancing treatments. The shift is greater under the Low-High order than under the High-Low. This is very significant for Ratings and the trends for Inches are in the same direction. Note that the significant heterogeneity of variance casts suspicion upon this one analysis.

For Discriminability treatments, all effects are convincingly absent. For Preshift Training, effects are generally absent, with the 5% level findings for Ratings being in the direction of greater shift under Short Phase 1 treatments, a finding similar to that of Johnson (1949). While this effect is primarily contributed by the one Low-High Short Rating group, which also accounts for the Phase 3 Training main effect, slight trends are in the same direction for High-Low Ratings, and also for the two Inches comparisons.

Possible Interpretations.—To return to the initial findings presented, our preferred interpretation for the striking discrepancies between the Phase 1 and Phase 3 analyses is this: The symmetrical and reversible findings for Inches represent the operation of a single perceptual contrast process. This process is also present, and reversible, for the Ratings judgments, but in addition, the Phase 1 Ratings data contain a second process of defining perceptual equivalents for novel judgmental terms. That this operates in an adaptation level (Helson, 1947, 1948, 1959) fashion is indicated in Table 1 by the closeness with which the mean ratings of *all* stimuli approach the designated "average" value of 100 in all treatments. Of the Phase 1 effects of context on Ratings (of the tracer Stimulus 7) about one-half is attributed to this adaptation-level definitional process, the other half to the perceptual contrast effect. This definitional process is deemed irreversible without explicit retraining involving instructional feedback. This irreversibility is particularly strong under our instructions, "always use the same number for the same length of line," but would probably be so under usual single stimuli instructions, except in so far as "vocabulary exhaustion" (Campbell, Lewis, & Hunt, 1958) forced redefinition. In Phase 3, the experimentally reversed perceptual contrast effect and the perseverating definitional contrast effects tend to cancel each other just as they augmented each other in Phase 1.

This interpretation leads to the expectation that the magnitude of the Phase 1-Phase 3 shift for both Inches and Ratings should be the same, inasmuch as both are primarily affected by perceptual contrast effect, resulting from the experimental context reversal. While our choice of a common graphic scale for Fig. 1 and 2 was a somewhat unsatisfactory compromise, the method of selecting it was independent of the Phase 1-Phase 3 shifts. Hence, the degree to which the shift magnitudes turn out to be similar for the two judgmental languages encourages this interpretation. Note particularly in Fig. 1 the parallel course which the two High-Low lines take, and similarly the two Low-High lines. Graphic superposition (not shown here) shows them to be practically indistinguishable. We anticipate that were a more unequivocable psychophysical scale equivalence to be achieved, as through a study in which all 13 stimuli occur in each "trial" with an unchanging "phase," the magnitudes of our Phase 1-Phase 3 shifts would not significantly differ when translated into the common metric.

Some other indirect evidence for the equivalence of shifts under the two conditions comes from the t ratios for shift, computed by pooling all Discriminability and Preshift Training conditions. For Ratings and Inches, respectively, these were 11.92 and 11.33 for Low-High, and 10.95 and 11.34 for High-Low ($df = 79$, $P < .001$, each case). If the t ratios can be used as comparable measures of magnitude of effects (since the number of *Ss* are equal), the effects are strikingly equivalent, with not the slightest trend

toward statistical significance of difference. This as evidence for equivalence is of course quite inferential. However, the mode of inference is quite independent of that based upon the conversion formula employed in the graphs. The approximate agreement of these two independent forms of inference is taken as encouraging the retention of the hypothesized two-process interpretation.

The temporal weighting of experience.— The attractiveness of the two-process interpretation may be increased by considering an alternative uniprocess theory which would presume quite different weighting parameters in the temporal accumulation of experience under the two conditions. Three temporal weighting outcomes can be considered: a primacy effect, in which the first visual experiences would have greater effect than later ones in determining the adaptation level at any given time; a recency effect, in which the adaption level would be predominantly determined by very recent experience; and an equal temporal weighting outcome, in which all experiences (within the experimental session at least) would be equally weighted. Inches judgments, as seen in Fig. 1 and 2, clearly show a strong recency effect. The reversibility is almost complete, the two curves cross almost at the center of the transition phase. Apparently only the current trial stimuli or the ones just preceding are having much influence on the adaptation level against which the tracer stimulus is being judged.

In contrast, the Ratings fit perfectly a model of equal temporal weighting. Under this assumption, the judgments of Stimulus 7 for the High-Low and the Low-High groups should converge at that trial at which the total of experimental stimuli presented was the same, for at that point, each would be judged against the same adaptation level. For the Short treatment, this would be at Trial 14, i.e., when the number of Phase 1 and Phase 3 trials were equal. This is exactly what happens. For the Long groups, this would happen at a projected Trial 33. From Fig. 1 it seems plausible that this might occur although a slight primacy effect is also plausible.

While each of these uniprocess interpretations is plausible, it does not seem reasonable that so small a difference in conditions as the reporting language could make such a fundamental difference in the perceptual and memory processes as to provide an equal weighting in one instance and a recency weighting in another. Furthermore, a time-specific recency effect is probably the general rule for all nervous system processes, both end-organ adaptation and memory. For example, end-organ adaptation as involved in judgments of temperature, salinity, brightness, etc. all show recency effects so dominant that the primacy-recency problem, or the temporal weighting problem in any form, is almost never raised. In judgment studies, Parducci (1954, 1959) and Campbell, Lewis, and Hunt (1958) have found recency effects, and recency effects are to be inferred from Johnson's (1949) data. Recency effects likewise seem universal in learning, even if concomitant with and occasionally disguised by primacy effects (e.g., Miller & Campbell, 1959).

Accepting the dual-process interpretation would lead to the conclusion that inferences as to temporal weighting made from single-reversal studies using the method of single stimuli in its usual forms are erroneous. These would include Helson's early experiment (1947, pp. 16–17) used to justify equal temporal weighting, and the studies of Campbell, Hunt, and Lewis (1957, 1958) and Segall (1959). The two-process interpretation would predict that recency weighting is more general and that apparent equal weighting or primacy effects are a function of the semantic rather than perceptual effects. The evidence assembled here is not crucial for the two-process interpretation. A more definitive approach would be multiple reversal studies of a Low-High-Low-High nature. With increasing numbers of reversals of context, the semantic factors should decrease in salience, allow-

ing the perceptual aspect to be manifest more directly.

Summary

The effects of divergent and shifting stimulus contexts upon the judgment of a common tracer stimulus were measured for the task of judging lengths of lines projected on a screen. For judgments made in inches, a reversible perceptual-contrast effect was demonstrated. Judgments made with an uncrowded numerical rating scale showed in addition a response-definitional contrast effect augmenting the apparent contrast effect in the initial phase and persisting into subsequent phases. These findings cast doubt upon previous estimations of the appropriate temporal weighting of stimuli in contributing to an adaptation level. Effects of stimulus discriminability and preshift training are also examined.

References

Bradley, R. A., & Schumann, D. E. The comparison of the sensitivities of similar experiments: Theory. *Ann. math. Statist.*, 1957, **28,** 902–920.

Campbell, D. T. Factors relevant to the validity of experiments in social settings. *Psychol. Bull.*, 1957, **54,** 297–312.

Campbell, D. T., Hunt, W. A., & Lewis, N. A. The effect of assimilation and contrast in judgment of clinical materials. *Amer. J. Psychol.*, 1957, **70,** 347–360.

Campbell, D. T., Hunt, W. A., & Lewis, N. A. The relative susceptibility of two rating scales to disturbances resulting from shifts in stimulus context. *J. appl. Psychol.*, 1958, **42,** 213–217.

Campbell, D. T., Lewis, N. A., & Hunt, W. A. Context effects with judgment language that is absolute, extensive, and extra-experimentally anchored. *J. exp. Psychol.*, 1958, **55,** 220–228.

Helson, H. Adaptation-level as a frame of reference for the prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60,** 297–313.

Helson, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55,** 1–29.

Helson, H. Adaptation-level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

Johnson, D. M. Learning function for a change in scale of judgment. *J. exp. Psychol.*, 1949, **39,** 851–860.

Krantz, D. L. Stimulus and response relationships in absolute judgment. Unpublished master's thesis, Northwestern University, 1959.

Miller, N., & Campbell, D. T. Recency and primacy in persuasion as a function of the timing of speeches and measurements. *J. abnorm. soc. Psychol.*, 1959, **59,** 1–9.

Parducci, A. Learning variables in the judgment of single stimuli. *J. exp. Psychol.*, 1954, **48,** 24–30.

Parducci, A. An adaptation-level analysis of ordinal effects in judgment. *J. exp. Psychol.*, 1959, **58,** 239–246.

Segall, M. H. The effect of attitude and experience on judgment of controversial statements. *J. abnorm. soc. Psychol.*, 1959, **58,** 61–69.

Stevens, S. S. Adaptation level vs. the relativity of judgment. *Amer. J. Psychol.*, 1958, **71,** 633–647.

(Received April 30, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 43–50

EFFECTS OF FOREPERIOD, FOREPERIOD VARIABILITY, AND PROBABILITY OF STIMULUS OCCURRENCE ON SIMPLE REACTION TIME[1]

D. H. DRAZIN[2]

Oxford University

Only recently has any systematic attempt been made to elucidate the relationship between the simple reaction time (RT), the foreperiod of the reaction (interval between warning signal and stimulus), and the variability of the foreperiod. Klemmer (1956) found that when the foreperiod was varied at random within a rectangular frequency distribution, mean RT increased as the mean foreperiod and range of foreperiods were separately increased. In a further study, employing normal distributions of foreperiods, Klemmer (1957) adduced evidence in support of the hypothesis that the mean RT varies in direct proportion to *S*'s uncertainty as to the timing of the stimulus, this uncertainty arising from two sources, variable error in *S*'s time-keeping and objective foreperiod variability. He estimated the former component on the basis of "prediction" tests, in which *S* was required to reproduce intervals equal to the mean foreperiods of matched simple reaction tasks. Corresponding foreperiod and prediction variances were summed and a conversion made to information measure (bits) relative to the uncertainty of the individual *S*'s predictions of a 1-sec. interval.

Although Klemmer's "informational hypothesis" offers an explanation of the relationship between mean RT and the parameters of the foreperiod frequency distribution, it provides no account of the relationship between RT and the foreperiod of the individual reaction (immediate foreperiod). While it must be conceded that Klemmer (1956) found no evidence of such an effect, contrary findings (Breitweiser, 1911; Karlin, 1959; Mowrer, 1940; Telford, 1931; Woodrow, 1914) of some consistency are on record and evidently require some explanation.

The present experiments, employing continuous rectangular foreperiod distributions, gave rise to pronounced RT-foreperiod curves comparable to those previously reported. It is shown that the form of the RT-foreperiod curve is systematically dependent on the parameters of the foreperiod frequency distribution. The effect of probability of stimulus occurrence was also examined. This variable is of special interest in connection with the common practice of administering a small proportion of check trials (wherein the stimulus does not occur) so as to minimize the frequency of anticipatory reactions.

METHOD

Apparatus.—The *S* received a 50-cps warning tone, which also served to mask extraneous noise in and about the laboratory, through close-fitting headphones. The stimulus was the onset of a Hivac CCIL neon lamp which *S* observed against a black background at a distance of 18 in. The warning

[1] Based on a thesis submitted in partial fulfilment of the requirements for the DPhil degree at Oxford University. The author wishes to express his gratitude to his supervisor, H. Kay, for advice and encouragement, and to the Department of Scientific and Industrial Research for financial support.

[2] Now at the RAF Institute of Aviation Medicine, Farnborough, England.

tone and stimulus remained on until *S* had responded by pressing a reaction key with the preferred index finger. The time recording equipment was stopped when a force of 120 gm. was applied to the key.

Foreperiod intervals were produced at random within a continuous rectangular distribution, for preset values of minimum foreperiod and range of foreperiods, by a system of relays and condenser-resistance time-delays. One time-delay, preset to the required range of foreperiods, cycled continuously, closing a relay momentarily at regular intervals. Another time-delay, initiated by *E*, closed a self-latching relay after an interval preset to the required minimum foreperiod. The simultaneous closure of both relays triggered a final self-latching relay which switched on the stimulus lamp. In order to ensure that *E* started the sequence of events in random phase with the cycling timer, all relays were silenced and care was taken to render the variability of the intertrial interval large in relation to that of the longest range of foreperiods employed. These precautions succeeded in producing foreperiod distributions which were not to inspection discriminable from rectangular.

The occurrence or nonoccurrence of the stimulus on a given trial was determined by the terminal state of a uniselector which rotated for a random period (the foreperiod) and which was so connected that a circuit between a selected set of wipers and the central pole of the selector switch was made on random trials according to a preset probability.

Individual foreperiods and RTs were recorded as cycle counts of a 1000 cps quartz crystal oscillator on decade counters.

TABLE 1

DESCRIPTION OF CONDITIONS: EXP. I AND II

Experiment I[a]			Experiment II[b]	
Cond.	Range of Foreperiods (Sec.)	Probability of Stimulus Occurrence	Cond.	Minimum Foreperiod (Sec.)
A	2.0	1.0	P	2.0
B	2.0	.9	Q	1.0
C	2.0	.5	R	0.5
D	1.0	1.0	S	0.25
E	1.0	.9	T	0.125
F	1.0	.5		
G	0.5	1.0		
H	0.5	.9		
J	0.5	.5		

[a] Mean foreperiod = 1.5 sec.
[b] Range of foreperiods = 1.0 sec.; probability of stimulus occurrence = 1.0.

Procedure.—Two separate experiments are reported. The conditions are described in Table 1. In Exp. I, mean foreperiod was held constant throughout at 1.5 sec., range of foreperiods and probability of stimulus occurrence serving as variables in a 3×3 factorial design.

The *S*s attended 5 practice and 12 test sessions, each of 50 min., over a period of three weeks. The practice sessions, comprising 30 trials in each condition, were run daily during the first week, the conditions being held in the same order every day.

The test sessions were run daily during the following two weeks. All probabilities of stimulus occurrence were always given in immediate succession for each range of foreperiods; though sequential effects of range, and of probability within each level of range, were completely counterbalanced within each set of six sessions. Ten warm-up trials were held prior to the first test run at each range, every run consisting of 30 trials as in the practice sessions.

In all sessions *S* was informed of the conditions of each ensuing run. He was instructed to respond as soon as possible after seeing the stimulus, while taking care not to respond in anticipation. In the practice periods, he was cautioned after responding prematurely. He was informed of his RT after each reaction and of his mean score in every condition after each session. A rest interval of 2 min. was allowed between runs. The experiment was held in a dimly lit room. The mean intertrial interval was about 10 sec.

In Exp. II, range of foreperiods was held constant at 1.0 sec. and probability of stimulus occurrence at 1.0, minimum foreperiod being held at five levels differing by factors of two. The *S*s attended 5 practice and 10 test sessions, each of 30 minutes, over a period of three weeks. In the practice sessions, conditions were given in the same order every day. In the experiment proper, order effects were counterbalanced within each set of five sessions. Each run consisted of 50 trials.

Subjects.—The *S*s were three University students aged between 19 and 26. Two *S*s served in each experiment.

RESULTS[3]

Experiment I.—The data were treated as follows. Taking *S*s and

[3] Complete data, consisting of mean RTs, *SD*s, and *N*s for all points in Fig. 1 and 2, unweighted mean RTs, and frequencies of anticipatory reactions, have been deposited with the American Documentation Institute.

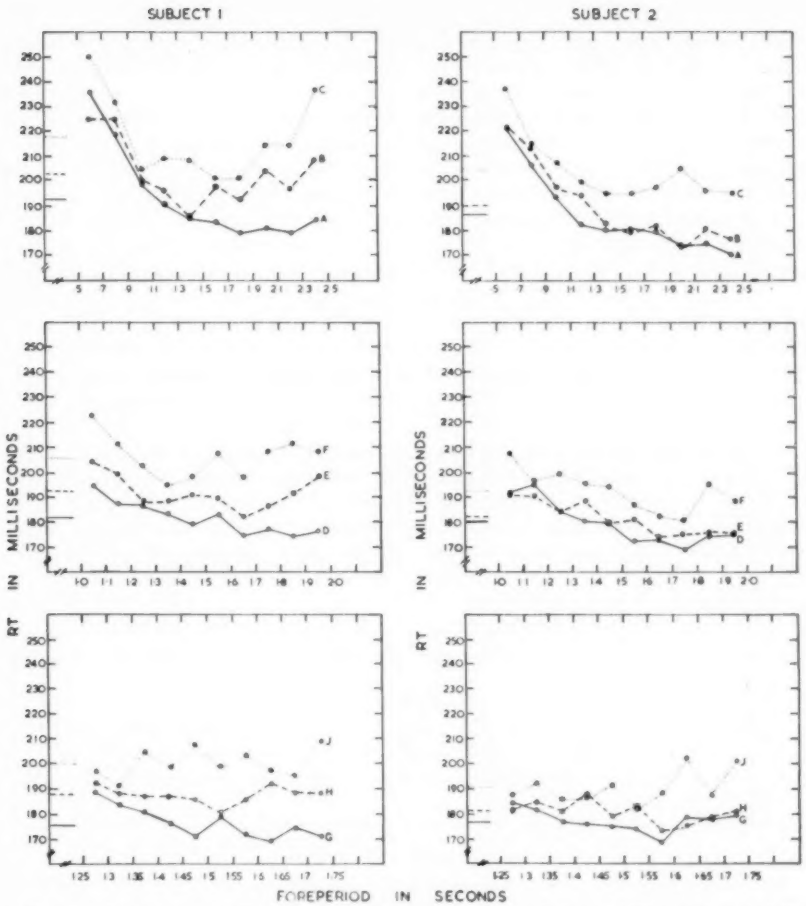

FIG. 1. Experiment 1: The influence of foreperiod, foreperiod variability, and probability of stimulus occurrence upon RT. (RT-foreperiod curves for individual conditions and *Ss*—see Table 1. Overall unweighted mean RTs indicated on the ordinate by horizontal straight lines.)

conditions separately, the range of foreperiods was divided into 10 equal sections and the mean of RTs associated through their foreperiods with each section was plotted against the expected section mean. The resultant RT-foreperiod curves are illustrated in Fig. 1, the horizontal lines on the ordinate indicating the overall mean RTs. False reactions (occurring before the stimulus) and premature reactions (occurring after, but presumed to have been initiated in the central nervous system before, the

Order No. 6683 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

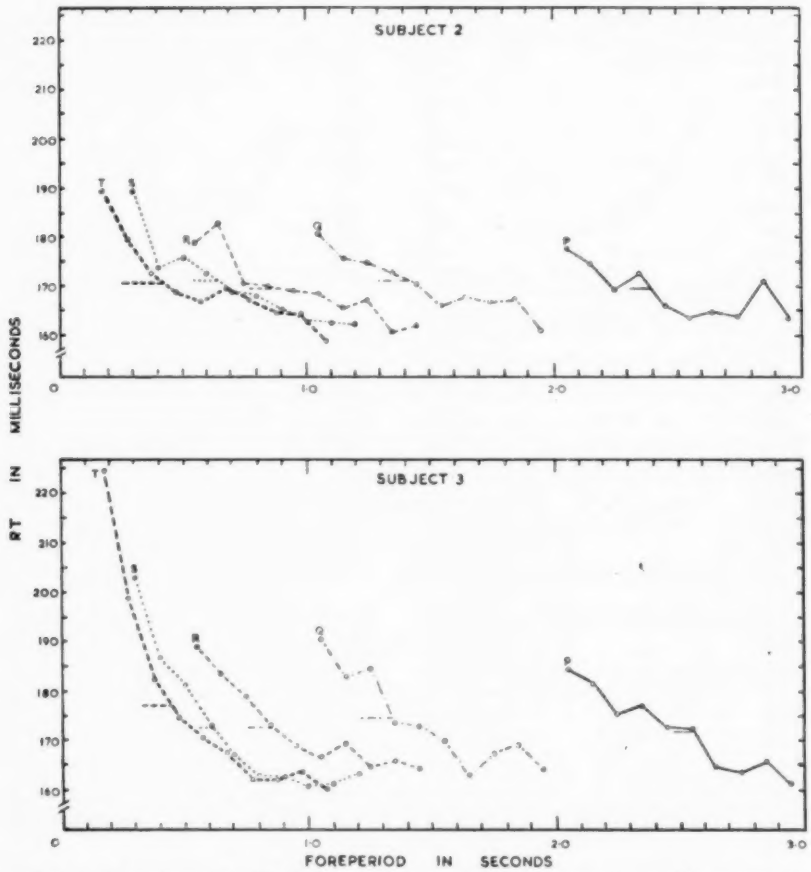

FIG. 2. Experiment II: The influence of foreperiod and foreperiod variability upon RT. (RT-foreperiod curves for individual conditions and *Ss*—see Table 1. Overall unweighted mean RTs indicated on regression curves by horizontal straight lines.)

stimulus) were excluded from this treatment. A reaction was discarded as premature if *S* judged it to be so or if the recorded RT was less than 100 msec. Since false and premature responses accounted for only 1.5% of trials, it is concluded that anticipation was a negligible source of error.

Analysis of variance of raw RTs was performed in order to determine whether the illustrated RT-foreperiod curves were compatible, at the 2.5% confidence level, with the hypothesis of simple regression.[4] Significant deviations from simple regression were obtained in all conditions except E, H, and J (S_1) and G, H, and J (S_2).

It may be seen that overall mean

[4] Slight heterogeneity of variance was noted but, in view of the different and complex relationships which appeared to hold between the cell means and variances in the several conditions of the experiment, it was decided not to attempt a transformation. Allowing for this source of error, the results of the analyses may safely be regarded as significant at the 5% level.

RT is an increasing function of the range of foreperiods, the mean foreperiod being held constant. Further, overall mean RT varies inversely with the probability of stimulus occurrence.

The effect of immediate foreperiod is pronounced for ranges of 1.0 and 2.0 sec., but generally nonsignificant for the 0.5-sec. range. The form of the RT-foreperiod curve varies with the probability of stimulus occurrence and also between *S*s. With unit probability of stimulus occurrence, RT decreases as a negatively accelerated function of immediate foreperiod, this relationship being most marked in Cond. A (2.0-sec. range of foreperiods). With the 2.0-sec. range of foreperiods, S_1's results show an increasingly steep positive terminal gradient with decreasing probabilities of stimulus occurrence. There is no indication of such a gradient in the results of S_2.

Experiment II.—The data of Exp. II were treated in the same way as those of Exp I. The resultant RT-foreperiod curves and overall mean RTs are illustrated in Fig. 2. Since false and premature responses accounted for only 1.2% of trials, it is concluded that anticipation was a negligible source of error. Analysis of variance of raw RTs showed that all RT-foreperiod curves were incompatible with the hypothesis of simple regression at the 0.1% confidence level.

It is evident that overall mean RT is not a critical function of minimum foreperiod when the range of foreperiods is held constant. As in Exp. I, the effect of immediate foreperiod is marked, long RTs again occurring with relatively short foreperiods. As the minimum foreperiod increases, the initial gradient of the RT-foreperiod curve becomes less steep and the curve as a whole more linear. Inter-*S* differences are again apparent: whereas the forms of the two *S*s' curves are broadly similar, those of S_3

TABLE 2

SIMPLE AND INTERACTION EFFECTS (IN MSEC.) OF IMMEDIATE FOREPERIOD (Fp), PRECEDING FOREPERIOD (Fp_{-1}), AND SECOND PRECEDING FOREPERIOD (Fp_{-2})

Effect	Condition				
	P	Q	R	S	T
Subject 2					
Fp	−6.82**	−9.37**	−9.44**	−12.19**	−10.12**
Fp_{-1}	−0.01	4.87**	3.47*	2.99*	1.18
Fp_{-2}	−0.53	1.11	−1.28	3.45*	0.17
Fp × Fp_{-1}	0.13	−1.91	−1.25	−1.07	1.85
Fp × Fp_{-2}	−1.21	1.69	0.37	2.54	1.17
Fp_{-1} × Fp_{-2}	−0.09	−2.61	0.63	1.01	−0.76
Fp × Fp_{-1} × Fp_{-2}	−1.44	−0.24	1.97	−1.13	−1.66
Estimated *SE*	1.35	1.33	1.64	1.37	1.52
Subject 3					
Fp	−13.18**	−13.93**	−13.83**	−20.96**	−27.76**
Fp_{-1}	4.14*	5.32*	7.40**	5.45**	0.60
Fp_{-2}	−0.68	1.96	3.52*	4.00*	2.03
Fp × Fp_{-1}	−1.21	−5.86**	−4.31*	−0.44	0.58
Fp × Fp_{-2}	−0.00	2.00	−0.01	−3.02	1.06
Fp_{-1} × Fp_{-2}	−4.54*	−2.29	1.53	2.93	−0.74
Fp × Fp_{-1} × Fp_{-2}	0.33	−0.05	0.09	−0.41	2.72
Estimated *SE*	1.75	1.72	1.39	1.66	1.97

* $P < .05$.
** $P < .01$.

show a more marked foreperiod effect.

The *SD*s about values plotted in Fig. 2 appear to vary systematically with the length of foreperiod relative to the range of foreperiods (see supplementary data). The results of both *S*s exhibit a similar trend in each condition: the scatter of RTs is greatest at relatively short intervals, decreases steeply to a minimum at about the mean interval, and then increases gradually towards the longest interval.

Previous workers (Karlin, 1959; Klemmer, 1956; Woodrow, 1914) have shown that the RT on any given trial is influenced to some extent by the lengths of preceding foreperiods relative to the mean foreperiod, longer RTs occurring when the foreperiod of the preceding trial is relatively long. Klemmer (1956) sought but failed to find any effects due to the foreperiod of the second preceding trial. The results of an analysis of the effects of the immediate foreperiod (Fp), the preceding foreperiod (Fp_{-1}), and the second preceding foreperiod (Fp_{-2}), based on the data of Exp. II, are presented in Table 2. Taking *S*s and conditions separately, all RTs, other than those recorded on the first and second trials of a run, were entered in a $2 \times 2 \times 2$ classification according to whether Fp, Fp_{-1}, and Fp_{-2} were separately greater or less than the expected mean foreperiod. The magnitudes of all main and interaction effects were estimated by taking sums and differences of means in the usual way. Significance was determined on the basis of estimated standard errors.

The fact that the main effects of Fp are all negative and highly significant indicates that short immediate foreperiods gave rise to long RTs, merely confirming the finding established by a more precise analysis of variance that the RT-foreperiod curves depart significantly from simple regression.

There would appear to be a discontinuity in the influence of preceding foreperiods between minimum foreperiods of 0.125 and 0.25 sec. In Cond. T (minimum foreperiod = 0.125 sec.) only the effect of Fp is significant, whereas several effects are significant in other conditions.

Discounting Cond. T, long RTs tend to follow trials with short foreperiods and vice versa, as shown by the fact that the main effects of Fp_{-1} are mostly positive and significant. There is a suggestion that Fp_{-2} has an influence similar to but weaker than that of Fp_{-1}. This suggestion is confirmed for S_3, the mean effect of Fp_{-2}, averaged over Cond. P, Q, R, and S, exceeding the compound *SE* by a factor of 2.70 ($P < .01$).

That the Fp $\times$ Fp_{-1} interactions are mostly negative suggests that the influence of the immediate foreperiod is greatest following trials with short foreperiods. This is confirmed for S_3, the mean value of this effect, averaged over Cond. P, Q, R, and S, exceeding the compound *SE* by a factor of 3.62 ($P < .001$).

It may also be noted that the effects of preceding foreperiods tend to diminish as the minimum foreperiod increases beyond about 0.5 sec. The Fp $\times$ Fp_{-2}, $Fp_{-1} \times Fp_{-2}$, and Fp $\times$ $Fp_{-1} \times Fp_{-2}$ interactions are generally nonsignificant.

Discussion

The results of the present study tend to confirm Klemmer's (1956) finding that overall mean RT, irrespective of immediate foreperiod, is an increasing function of the range of foreperiods and that where the range is appreciable (say more than 0.5 sec.) overall mean RT is not a critical function of the mean foreperiod. It is notable that S_2's and S_3's overall mean RTs in the five conditions of Exp. II are narrowly scattered over ranges of about 2 and 5 msec., respectively. Similarly Klemmer, using

a 2.0-sec. range, obtained the same overall mean RT with minimum foreperiods of 0.25 and 3.25 sec.; with a minimum foreperiod of 6.25 sec. the overall mean RT was only about 13 msec. longer.

Klemmer's (1957) hypothesis that overall mean RT is a linear function of *S*'s uncertainty as to the timing of the stimulus is contraindicated by the present data. Woodrow (1930) found that, for intervals less than about 4 sec., the variability of temporal judgments (reproduction of temporal intervals) was roughly proportional to the time judged, the *SD* being about 10% of the standard interval. Thus, according to Klemmer, we should expect that, when range is held constant, overall mean RT would be an increasing function of the minimum foreperiod. It has been noted, however, that the results of Exp. II showed no trend in this direction.

Most of the RT-foreperiod curves obtained in the present study and many illustrated in the literature (e.g., Breitweiser, 1911; Karlin, 1959; Woodrow, 1914) exhibit marked initial gradients. Neither the characteristics nor even the existence of such gradients are explicable in terms of Klemmer's "informational hypothesis," since this only relates time uncertainty to the overall mean RT. Given the finding that RT is related to the immediate foreperiod, then it would seem that the characteristics of this relationship must receive prior explanation: a theory which could account for the entire RT-foreperiod curve would a fortiori account for the overall mean RT. That Klemmer failed to obtain evidence of an initial gradient in the RT-foreperiod relationship may possibly be due to a fundamental difference in experimental procedure. Klemmer used ranges composed of 11 equally spaced foreperiods delivered in random order (i.e., a discrete foreperiod frequency distribution bound by a rectangular envelope). For very short intervals, where the difference between alternative foreperiods (interforeperiod difference) must have been easily discriminable, *S* may have learned to "focus his attention" at or about the moments when there was a definite probability that the stimulus would occur, and by so doing he might have been able to respond more quickly (Mowrer, 1940). In the present study, however, foreperiod variability was continuous, so that such tactics would have been inappropriate. Concordant with this suggestion is the fact that in Klemmer's (1956, Fig. 1) variable foreperiod tests, where the interforeperiod difference was of the same order of magnitude as the minimum foreperiod, the mean RTs for the shortest intervals were below the overall mean RTs for their respective ranges but, where the interforeperiod difference was relatively small, the mean RTs for the first intervals were, as in the present study, somewhat above the general means.

Woodrow (1914) and Karlin (1959) have suggested that the RT-foreperiod relationship reflects progressive changes in *S*'s state of readiness. According to this view, the effects of preceding foreperiods are attributable to an adaptation process: on average, readiness reaches a peak or plateau at about the mean interval, but this peak or plateau tends to be attained earlier or later, according to the length of the preceding interval. The present results confirm the findings of Karlin (1959), Klemmer (1956), and Woodrow (1914) that RT varies as an increasing function of the preceding foreperiod. It has also been shown that RT varies in a like manner but to a less extent with the second preceding foreperiod. This suggests that the adaptation process which has been postulated is heavily biased in favor of recent intervals, but may be sensitive to quite remote ones.

The present finding that the scatter of RTs is greatest at relatively short intervals and least at intermediate ones may be interpreted as further support for the adaptation hypothesis. Readiness will be most variable at the minimum foreperiod, since on some occasions, following a series of relatively long intervals, the stimulus will occur at a point of minimum readiness, whereas on other occasions, following a series of relatively short intervals, the stimulus will occur at a point of maximum readi-

ness. At intermediate intervals, however, *S*'s state of readiness will rarely be much less than maximal and a smaller scatter of RTs might therefore be predicted.

It is interesting to note two areas in which the evidence concerning the effects of preceding foreperiods is in apparent conflict. First, while Karlin (1959) found that the effect of the immediate foreperiod was greatest following long preceding foreperiods, the present results show this effect to have been greatest following short preceding foreperiods. Further, it would seem that Klemmer (1956) found no significant trend in either direction.

Second, whereas Karlin obtained anomalous but significant effects due to preceding foreperiods in conditions where the minimum foreperiod was less than 0.8 sec., the present results indicate no significant effects of previous foreperiods with very short minimum foreperiods.

It is to be hoped that discrepancies such as these will be resolved by further studies, so that the characteristics of the postulated adaptation mechanism can be specified more clearly. It might then be possible to predict the form of RT-foreperiod curve which would result from the use of any random distribution of foreperiods. The mean RT for any foreperiod could be calculated simply by weighting the RTs characteristic of all possible patterns of preceding foreperiods by their expected frequencies within the distribution in question.

In Exp. I it was found that the RT-foreperiod relationship was dependent on the probability of stimulus occurrence, and that the form of this dependence was subject to sizable individual differences. It would be premature to speculate on the general significance of these results or their specific relevance to the adaptation hypothesis discussed above. At present it will suffice to indicate the necessity of controlling the frequency of check trials in investigations into the form of the RT-foreperiod relationship.

Summary

In two parallel experiments, the foreperiod of a simple reaction task was varied at random within a rectangular frequency distribution. In a total of 14 conditions, the minimum foreperiod, range of foreperiods, and probability of stimulus occurrence were systematically varied.

The results indicate that the RT-foreperiod relationship is subject to a range effect: in all conditions where the range of foreperiods exceeded 0.5 sec., RT tended to decrease initially as a negatively accelerated function of the foreperiod. The detailed form of this relationship was found to depend upon the minimum foreperiod, the range of foreperiods, and the probability of stimulus occurrence. The scatter of RTs was greatest for short foreperiods and least for foreperiods of intermediate length.

Reaction time was also found to vary with the foreperiods of the preceding reaction and, to a less extent, the second preceding reaction: in either case, long RTs tended to follow reactions with short foreperiods and vice versa. The effect of immediate foreperiod was most marked following reactions with short foreperiods.

The experimental findings are interpreted as evidence in favor of the view that RT reflects *S*'s state of readiness for the stimulus.

References

Breitwieser, J. V. Attention and movement in reaction time. *Arch. Psychol., NY*, 1911, **2**, No. 18.

Karlin, L. Reaction time as a function of foreperiod duration and variability. *J. exp. Psychol.*, 1959, **58**, 185–191.

Klemmer, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

Klemmer, E. T. Simple reaction time as a function of time uncertainty. *J. exp. Psychol.*, 1957, **54**, 195–200.

Mowrer, O. H. Preparatory set (expectancy): Some methods of measurement. *Psychol. Monogr.*, 1940, **52**(2, Whole No. 233).

Telford, C. W. The refractory phase of voluntary and associative responses. *J. exp. Psychol.*, 1931, **14**, 1–36.

Woodrow, H. The measurement of attention. *Psychol. Monogr.*, 1914, **17**(5, Whole No. 76).

Woodrow, H. The reproduction of temporal intervals. *J. exp. Psychol.*, 1930, **13**, 473–499.

(Received May 3, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 51-57

FREE RECALL AND ORDERING OF TRIGRAMS[1]

LEONARD M. HOROWITZ[2]

Johns Hopkins University

Some explanation is needed for the fact that intralist similarity can have either inhibitory or facilitative effects on the learning and retention of verbal material. The specific role of intralist similarity seems to depend upon the nature of *S*'s task, although the critical task variables are not known. The purpose of the present study is to examine some of these critical variables.

Underwood (1952, 1954) varied the similarity relationship between the items of a list learned by serial anticipation and concluded that "recall and intralist similarity are not systematically related" (1954, p. 160). To vary intralist similarity, Underwood used lists of consonant-vowel-consonant (CVC) trigrams which contained as many different letters as possible (low intralist similarity) and other lists in which certain letters occurred repeatedly throughout the syllables (high intralist similarity). More trials were needed to learn, and later to relearn, the high similarity lists to criterion. But the number of overt errors per trial during learning was not related to intralist similarity, nor was recall 24 hr. later.

Underwood defines intralist similarity as the extent to which the items of a list share letters in common with one another. Intralist similarity can also be expressed in terms of the sequential restrictions between the letters of the items. For example, particular groupings of two or more letters might occur with a high probability in the syllables of one list while these relationships could be left to chance in another list. The list in which the sequential dependencies depart markedly from randomness would be called highly redundant (Miller, 1958).

Miller (1958) set up explicit rules to generate redundant lists of items four to seven letters long. He also used lists in which the item letters were combined at random. When redundancy was introduced into Miller's lists, the items became more similar (by Underwood's definition) to one another. Miller has demonstrated that the redundant lists had greater intralist similarity by a task in which *Ss* search for particular items from a list of redundant items and from a list of random items. The median search time was considerably higher for the redundant lists.

Each of Miller's lists was presented in various orders, and after each presentation *Ss* were tested by the method of free recall. Recall of the redundant lists (high intralist similarity) was superior to that of the random lists, not only throughout original learning but also after the interpolation of another list. Since Underwood found that recall and intralist similarity are not systematically related, this outcome is rather surprising.

Aborn and Rubenstein (1952; Rubenstein & Aborn, 1954) have found essentially the same result as Miller with verbal material more like Underwood's. Whereas Miller's material contained four different consonants only, Aborn and Rubenstein used 16 CVC trigrams involving 19 different letters. The syllables were put together to form a list, but the lists ranged from a random assortment of trigrams to a highly redundant one ("highly organized," in their terms). As in Miller's data, the higher the intralist similarity, the greater the mean recall score.

The measure of learning in the Aborn and Rubenstein studies was more like Miller's method of free recall than Underwood's serial learning since, for the most part, the serial order of *S*'s recall did not affect his score. If a trigram was correctly placed within a grouping of four items (whether the grouping itself was recalled early or late), it was scored as a correct item. It was also scored correct it if merely followed another trigram appropriately in recall.

Thus the principal difference between Underwood's study of serial anticipation and the studies by Miller and Aborn and Ruben-

[1] This paper is based upon a dissertation submitted in partial fulfillment of the requirements for the PhD degree at Johns Hopkins University. It reports work done while the author was a Fulbright Fellow at University College, London, and a Social Science Research Council Fellow at Johns Hopkins University. The author is indebted to James Deese and A. R. Jonckheere for their enthusiastic advice and encouragement.

[2] Now at Stanford University.

stein is not in the concept needed to describe the interitem relationship (similarity), but in the measures of learning used. The major difference between serial learning and free recall is that serial learning requires a unique ordering of the items, while free recall does not.

The closest experiments to a study of the ordering process are discrimination studies like that of Fitts, Weinstein, Rappaport, Anderson, and Leonard (1956). Discrimination, like the ordering part of serial learning, places a premium on the uniqueness of items in the test material. Fitts et al. found that randomly generated geometrical figures can be recognized more rapidly than statistically constrained figures. In other words, redundancy is a hindrance to visual pattern recognition. It would not be surprising, then, to discover that redundancy is a hindrance to ordering as well.

The evidence suggests that high intralist similarity may facilitate free recall learning and hinder ordering. When ordering and free recall are combined in the serial anticipation method, serial learning could appear unrelated to intralist similarity. In view of this possibility, the first step in an analysis of intralist similarity would be to study the effects of such similarity upon free recall and ordering separately. That is the purpose of the present experiment.

METHOD

Materials

Two lists of 3-consonant (CCC) trigrams were used. One list contained fewer different letters than the other and therefore had greater intralist similarity. The high-similarity list was composed of four letters and is referred to as L4. It made use of F, S, V, and X to make a total of 12 items. These items are VXF, XVS, FSV, SFX, XFV, FXS, VSX, SVF, VFS, FVX, SXV, and XSF. Each of the four letters occurs three times as the initial letter of a syllable, three times as the middle letter, and three times as the final letter. A single letter occurs only once in an item.

This list contains half of the 24 possible three-letter combinations. Moreover, any one item in the group can be transformed into another item by applying the following rule: Replace any one letter of a three-letter combination by the fourth letter and interchange the remaining two letters of that item. For example, the first item of the list is VXF. If the F of this item is replaced by an S, and the V and X are interchanged, the new item is XVS, the second item on the list. Application of this rule yields precisely the above 12 items and no others. The rule does not generate the reversals of the items, though. (FXV is a reversal of VXF and does not occur.)

Each pair of letters occurs twice—once as the first two letters of a syllable and once as the last two letters of a syllable. In both cases the pair occurs in combination with the same third letter. (FX, for example, occurs in the item FXS as well as SFX.)

The other list is called L12 because it contains 12 different letters. The L12 items are: QSZ, KXW, MRV, WQJ, VKH, ZMQ, HWS, FVX, JZR, XHF, SJM, and RFK. This list has the following properties: (*a*) Each letter occurs once in each position; (*b*) a particular pair of letters occurs in only one item; and, of course, (*c*) all possible pairs or triplets of the letters do not appear in the list.

Methods of Presentation

Each list was presented to two groups of *Ss*. One group was required to learn the items by the method of free recall. The other was required to learn the *order* of the items, but not the items themselves.

Free recall.—L4 was presented one trigram at a time, on 4 × 5 in. flash cards. Each trigram was shown for 4 sec. After the *Ss* had seen the entire list, they were instructed to record on their answer sheets, in any order, as many of the items as they could remember; guessing was permitted. The answers were collected, and the items were presented again in a different order. This was done in order to minimize self-instruction for ordering among the *Ss*. A second group of *Ss* was presented with L12 in the same way.

Ordering.—L4 was presented to a group of *Ss* in a standard order of presentation. The order was the order in which the items are listed above. After each presentation, *Ss* were supplied with a packet of 12 1 × 1¾ in. slips of paper. Each slip of paper had on it one of the L4 items. The *Ss* were instructed to re-arrange these 12 slips so that the order corresponded to the order in which the flash cards had been presented. (The order of the items in the packet was different on each of the 10 trials.)

A fourth group ordered the L12 items under the same procedure. The standard order of the L12 items was the order in which they are listed above.

TABLE 1

MEAN CORRECT RECALL SCORES AND MEAN TOTAL RESPONSE SCORES WITH L4 AND L12

Trial	Mean Number of Syllables Correctly Recalled		Mean Number of Responses Given	
	L4	L12	L4	L12
1	3.12	1.83	7.20	6.08
2	4.80	3.54	9.24	7.75
3	5.72	4.79	10.08	8.62
4	6.24	5.96	10.40	10.08
5	6.56	5.92	10.52	9.79
6	6.64	7.12	10.64	10.54
7	6.60	7.42	10.20	10.42
8	7.00	8.00	10.68	10.42
9	7.16	8.46	10.96	10.08
10	7.79	8.50	10.44	11.08

Subjects

Each group contained the children of a preformed class in the Sir William Collins School, a secondary modern school in London, England. These groups were comparable in age, IQ, and academic achievement, and *Ss* within each group were tested simultaneously. There were 24 and 25 *Ss* in the free recall groups, and 30 and 19 *Ss* in the ordering groups. (The groups were unequal in size since the classes were unequal in size, but this difference does not affect the statistical analysis of the data.) All *Ss* were male, 14 or 15 yr. old, and slightly above average in intelligence.

RESULTS AND DISCUSSION

The most general result is that free recall for L4 is higher than for L12 early in learning (although the superiority is lost on later trials), while L12 is consistently easier to order than L4. The detailed results for the two methods are reported below.

Free Recall

Correct responses.—Table 1 shows the mean number of syllables correctly recalled on each trial and the mean number of responses (correct and incorrect) that *Ss* produced. L4 is superior to L12 in correct recall for Trials 1 through 5. Tested by the Mann-Whitney U test, the initial difference is significant at $P < .001$. Table 2 summarizes the analysis of variance of the correct recall scores. The Trial × List interaction is significant at $P < .001$, although on the average, the difference between the two lists is not significant.

The superiority of L4 in free recall is probably the result of chance guessing. Each L4 *S* must have learned all the L4 letters quite early since each *S* used all four letters in his responses on every trial. Furthermore, each response was a three-letter combination. If *S* randomly combined the four L4 letters into three-letter combinations, the probability of his producing a correct item is $\frac{1}{2}$. Table 1 shows that the mean correct score for L4 on the early trials is about half the mean total response score. Thus, the advantage of the high-similarity list is that it enables *Ss* to correctly guess a large number of items.

Just why L4 lost its superiority, though, is not completely clear. This loss of superiority may be due in part to the extreme similarity of the L4 items to each other. In the later stages of free recall learning, two or more very similar items may general-

TABLE 2

ANALYSIS OF VARIANCE OF CORRECT RECALL SCORES

Source of Variance	*df*	*MS*	*F*
Between *Ss*	48		
Lists	1	0.012	—
Error (b)	47	21.365	
Within *Ss*	441		
Trials	9	151.269	50.79*
Lists × Trials	9	11.797	3.96*
Error (w)	423	2.978	
Total	489		

* $P < .001$.

ize to the same erroneous response at the expense of a correct item. Such generalization may be difficult to overcome so that further improvement is hindered. This issue needs to be studied further with more levels of intralist similarity.

Intrusions.—There were 198 letter-intrusions throughout the responses of L4. E, U, and Y made up 72% of these intrusions. The responses of L12 contained 140 intrusions, and all 14 possible intrusions occurred. Since *S*'s recall was handwritten, the absolute number of intrusions may be slightly biased by possible errors in scoring. Therefore, a statistic summarizing the variability of intrusions is probably more appropriate here: the uncertainty in the L12 intrusions was 3.36 bits, while that in the L4 intrusions was 2.14 bits. Thus, L12 led to a greater variety of intrusions than L4. One reason for this difference in distribution is that the 12 letters of L12 evoked more different verbal habits than did the 4 letters of L4.

Ordering

Kendall's *P* (1955, pp. 3–7) was selected as the best measure of "goodness of ordering." All pairs of items are considered one by one in *S*'s returned packet; and whenever two items are correctly placed relative to one another, +1 is scored for that pair. Kendall's *P* may be considered the total number of +1's in a particular ordering.[3] $P = 66$ and $P = 0$ indicate perfect ordering and perfect reverse ordering, respectively, of 12 items. $P = 33$ is the mean chance expectation and indicates no relationship between the correct order and *S*'s order. The *SD* of *P* values for chance orderings of 12 items is 7.3.

TABLE 3

MEAN ORDERING SCORES AND MEAN CORRECT PLACEMENT SCORES WITH L4 AND L12

Trial	Mean Ordering Score (Kendall's *P*)		Mean Correct Placements	
	L4	L12	L4	L12
1	35.48	52.33	1.44	4.70
2	41.21	55.14	3.41	5.79
3	43.70	59.07	2.78	7.73
4	45.45	60.30	3.53	8.23
5	46.45	61.63	3.94	8.73
6	43.40	63.63	3.11	10.57
7	47.27	64.70	4.00	10.57
8	49.77	65.03	5.71	11.20
9	48.43	65.00	5.17	11.00
10	56.22	65.23	6.61	11.20

Table 3 gives the mean *P* score on each trial. It is clear that the L12 items are more distinctive and easier to order. The L12 group starts off at a higher origin, reaches a higher asymptote, and improves more smoothly.

Representative serial position curves (Trials 1, 5, and 9) are shown in Fig. 1. The ordinate of Fig. 1 is the percentage of *S*s who placed a particular item in its correct position—irrespective of the placement of the other items. L12 begins with a serial position effect on Trial 1 and improves to nearly perfect ordering (Trial 9) where the serial position effect just about disappears. A serial position effect for L4 does not really emerge until quite late in learning. In fact, the later trials of L4 resemble most the earlier trials of L12. This can be seen from Table 4, which contains correlations of "percent correct" scores between L4 and L12, position by position, for representative trials.

The conclusion comes up again that ordering is retarded with L4. The high similarity of the items of

[3] Because *P* has this simple, direct interpretation, it seems better suited here than Kendall's tau. *P* is, however, directly related to tau. For an ordering of 12 items, $\text{tau} = \frac{P-33}{33}$.

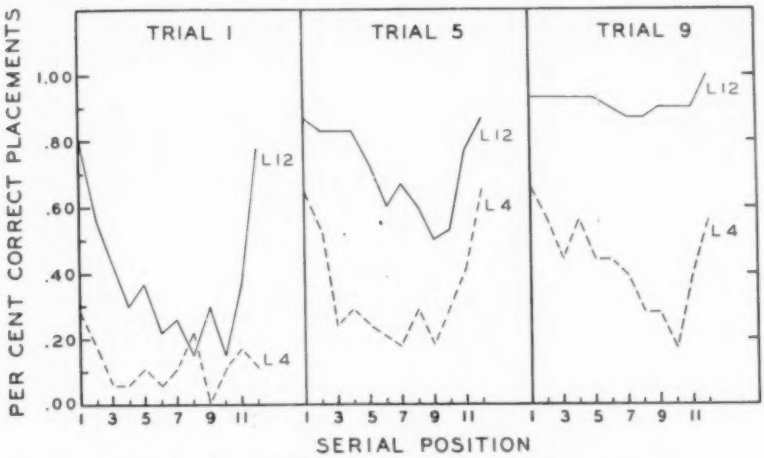

FIG. 1. Serial position curves on Trials 1, 5, and 9 for L4 and L12.

this list makes it extremely difficult for *Ss* to discriminate sufficiently between the items to learn a unique ordering of them.

Another way to look at the serial position data is to ignore relative ordering and consider the number of items that *Ss* placed in precisely the correct position on each trial. With this kind of measure—total number of correct placements—random arrangements of 12 items by a large population of *Ss* would give essentially a Poisson distribution of scores (Feller, 1957, pp. 97–99) with a mean of 1. It can further be shown that if the precise location of k items is memorized and the remaining items are assorted randomly, the best estimate of the mean correct placement score would be $k + 1$.

The mean correct placement scores are listed in Table 3. Prior to Trial 8, the mean for L4 does not rise above 4.0. If *S* memorized the first, second, and last items and let the remaining items fall randomly, the expected mean would be 4.0. It is not until the mean score rises above this point (Trial 8) that a real serial position effect emerges. Thus, it seems that with the highly similar items of L4, the items are not ordered relative to one another, but are memorized individually.

If seven items are ordered 1, 3, 4, 5, 6, 7, 2 the ordering score is $P = 16$ (out of a maximum of 21) and the correct placement score is 1 (out of a maximum of 7). Thus, excellent relative ordering can be obscured by the correct placement score. In order to show how the two measures are related in empirical data, correct placement scores have been plotted against ordering scores in Fig. 2.

TABLE 4

CORRELATIONS, POSITION BY POSITION, BETWEEN "PERCENT CORRECT" SCORES ON L4 AND L12

L12	L4		
	Trial 1	Trial 5	Trial 9
Trial 1	0.39	0.84	0.80
Trial 5	—	0.73	0.90
Trial 9	—	—	0.62

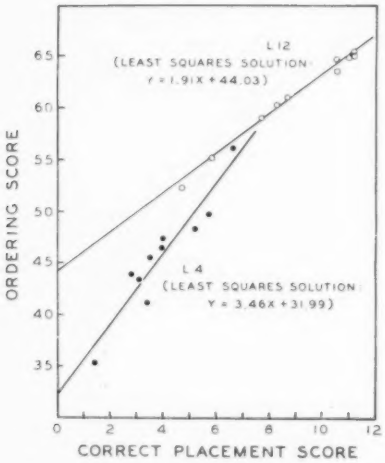

FIG. 2. Relationship between ordering scores and correct placement scores for L4 and L12.

For both lists the regression is linear, but the slopes differ considerably. If ordering improved as a function of correct placements only, about a three-point increase in ordering would occur to a one-point increase in correct placements. This is what happens with L4. When the correct placement score is 1, the ordering score is approximately 36. But the L12 ordering score starts considerably higher than one would predict from the 3:1 relationship. In fact, the L12 regression line suggests that had it been possible to measure performance early enough in learning, an ordering score of 46 would have corresponded to one correct placement. After *S* has reached a relatively good level of ordering, though, there is not much room for improvement, and the slope of the regression equation for L12 would have to be, and is, lower than that of L4.

These findings indicate the presence of an interaction between intralist similarity and *S*'s task. High intralist similarity facilitates free recall early in learning, but this superiority is lost on later trials. Low intralist similarity uniformly facilitates ordering.

When *S* learns by the method of serial anticipation, he has at least two tasks. He has to recall a set of responses (the items of the list) and he also has to order these responses. In view of the interaction between intralist similarity and *S*'s task, it is possible that, on the average, the importance of intralist similarity might be underestimated by the method of serial anticipation. Furthermore, this interaction might account for Underwood's (1954) conclusion that "recall (after 24 hr.) and intralist similarity are not systematically related" (p. 160).

A recent study by Underwood, Runquist, and Schulz (1959) indicates that paired-associate learning can be analyzed into two phases: (*a*) a response-learning phase, when responses become readily recallable, and (*b*) an associative phase, when each response is paired with a specific stimulus. One of their findings closely resembles a finding of this study: that high intralist similarity facilitates response learning. Since high intralist similarity hinders overall paired-associate learning (Feldman & Underwood, 1957), Underwood et al. (1959) inferred that the interference component of high intralist similarity occurs in the associative phase. In this respect the associative phase of paired-associate learning is like the ordering part of serial learning. This notion could be tested directly by teaching *S*s to match cards to appropriate stimuli without actually having them learn the responses per se. Thus, the method of the present study is directly applicable to the study of paired-associate learning.

The present results, furthermore, suggest the following way to maximize overall efficiency when *S* has to learn a number of lists by the serial anticipation method or the paired-associate method. The entire pool of items should first be grouped into lists of maximum intralist similarity for learning by the method of free recall. Then the items should be

regrouped into lists of minimum intralist similarity for ordering or for associative learning. If there is no negative transfer from free recall to ordering or to associative learning, serial learning or paired-associate learning should then proceed very rapidly. This analysis needs to be tested experimentally.

Summary

This study was designed to examine the joint influence of intralist similarity and the nature of *S*'s task on verbal learning. Two lists of 12 3-consonant trigrams were used. The trigrams of the high-similarity list contained only 4 different letters, while those of the low-similarity list contained 12 different letters. Two groups of *S*s recalled as many syllables as they could after each of 10 presentations, while two other groups learned the *order* of the syllables but not the syllables themselves.

The results indicate the presence of an interaction between intralist similarity and the nature of *S*'s task. High intralist similarity facilitates free recall early in learning, but this superiority is lost on later trials. Low intralist similarity, on the other hand, uniformly facilitates ordering. Since recall and intralist similarity have not been found to be systematically related when the method of serial anticipation was used, this interaction may be important in understanding the method of serial anticipation.

It was suggested that the method of this study might be applied to the study of paired-associate learning. The results also imply a way to maximize overall efficiency when *S* has to learn by the method of serial anticipation or by the method of paired-associates.

References

Aborn, M., & Rubenstein, H. Information theory and immediate recall. *J. exp. Psychol.*, 1952, **44,** 260–266.

Feldman, S. M., & Underwood, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53,** 11–15.

Feller, W. *An introduction to probability theory and its applications.* New York: Wiley, 1957.

Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, 1956, **51,** 1–11

Kendall, M. G. *Rank correlation methods.* London: Griffin, 1955.

Miller, G. A. Free recall of redundant strings of letters. *J. exp. Psychol.*, 1958, **56,** 485–491.

Rubenstein, H., & Aborn, M. Immediate recall as a function of degree of organization and length of study period. *J. exp. Psychol.*, 1954, **48,** 146–152.

Underwood, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity. *J. exp. Psychol.*, 1952, **44,** 80–87.

Underwood, B. J. Intralist similarity in verbal learning and retention. *Psychol. Rev.*, 1954, **61,** 160–166.

Underwood, B. J., Runquist, W. A., & Schulz, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58,** 70–78.

(Received May 4, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 58–61

FORWARD, BACKWARD, AND PSEUDOCONDITIONING OF THE GSR

R. A. CHAMPION AND J. E. JONES[1]

University of Sydney

For some time it has been customary to class forward and backward conditioning as fundamentally different processes and to regard the latter phenomenon as a form of pseudoconditioning (Underwood, 1949) or sensitization (Osgood, 1953) rather than true conditioning. While the evidence for this distinction is strong in the case of short-latency CRs (e.g., Spooner & Kellogg, 1947), there are theoretical reasons for believing that some true learning may occur when a backward conditioning procedure is used with a long-latency response (Mason, 1959). One reason for this expectation is that the presentation of the UCS prior to the CS by an appropriate time interval should allow greater contiguity of the CS and the UCR if the latter has a long latency. If it be supposed, however, that reinforcement is a necessary condition for learning as well as S-R contiguity then the backward conditioning procedure may mean greater delay of reinforcement and less propitious conditions for learning. Considering degree of S-R contiguity as the interval between the onset of the CS and the onset of the UCR, there may be greater contiguity in the backward procedure than in the forward procedure. On the other hand, assuming that proximity of reinforcement is the interval between the occurrence of the CR and the offset of the UCS, there may be greater proximity in the forward procedure than in the backward procedure. The optimum arrangement of stimuli for conditioning may be expected to vary with the latency of the response being conditioned, among other factors. In the present state of knowledge these optimum conditions cannot be accurately specified, it not being known, for example, that the reinforcement in classical aversive conditioning necessarily occurs with the offset of the noxious UCS, nor that any reinforcement at all is required in classical conditioning. As a first approximation the type of theory referred to above leads to the possibility of obtaining some degree of true backward conditioning of the GSR and the following data allow a test of this possibility. When the experiment was first designed it was assumed that a backward conditioning schedule would provide a valid control for pseudoconditioning, but this procedure produced evidence for true conditioning in the form of a clear and continued rise in performance and a group was therefore also tested with no short-term contiguity of CS and UCR, either forward or backward.

METHOD

Subjects.—The *S*s were 32 male and 28 female volunteers from undergraduate courses in psychology at the University of Sydney.

Apparatus.[2]—The *S* was seated in a sound-reduced room separate from that containing *E* and the main apparatus. The CS was a 2000-cps, 60-db. tone of 20 msec. duration presented through a speaker immediately behind *S*'s back. The ambient noise level

[1] Formerly Joan E. Mason and now at the Defence Research Medical Laboratory, Toronto, Canada.

[2] Part of the apparatus used in this study was provided by Commonwealth of Australia Research Grant No. 2002.

in *S*'s room was approximately 50 db., being contributed to by the sound of an electric circulation fan. The UCS was a spike-wave ac shock of 90 cps, 3.20 milliamp., and 600 msec. duration. Shock strength was measured by means of an oscilloscope with a 15,000-ohm resistor in place of *S*. The shock was delivered to *S* through copper electrodes in the form of spring clamps 15 mm. wide which almost encircled the third and fifth fingers of the right hand near the fingertips. The GSR was picked up through thimble-type electrodes of German silver 30 mm. in length on the second and fourth fingers of the right hand. The shock and GSR electrodes were used in conjunction with electrode jelly. The GSR was measured and amplified with an electronic dc amplifier of the bridge type supplying constant current of 100 microamp. to the electrodes and followed by an ink-writing recorder. The time intervals listed above were determined by electronic interval timers, but when shock was used in isolation as a stimulus its duration was determined manually by *E* to avoid any possibility of giving cues to *S* and approximated 600 msec.

Procedure.—Following the receipt of instructions each *S* was given five trials with the UCS alone; the first and second shocks were 1.60 and 2.27 milliamp., respectively, and the remainder were 3.20 milliamp. The intertrial interval depended on *S*'s skin resistance reaching a steady level and was at least 60 sec.; the interval rarely exceeded 75 sec. and averaged about 65 sec. No ready signals were given at any stage of the experiment.

After the five preliminary UCS trials, Group TS ($N = 20$) was put on a forward conditioning schedule with 500 msec. between CS onset and UCS onset. Four CS-alone trials (T) were interspersed with the training trials (TS) as tests of conditioning and three more UCS-alone trials (S) were also included. The order of trials was TS, T, S, TS, TS, T, TS, S, TS, S, T, TS, TS, T. Group ST ($N = 20$) was put on a backward conditioning schedule with 750 msec. between UCS onset and CS onset; the order of trials was the same as for Group TS, but UCS-CS trials (ST) replaced the CS-UCS trials (TS).

After the data had been obtained from these two groups a third group of 20 *S*s was tested with a similar series of stimuli but with discrete presentations of tone and shock (Group T & S). The *S*s in this group received as many tones and shocks as the members of Groups TS and ST, but the stimuli were given in a different random order for each *S* and the minimum intertrial interval of 60 sec.

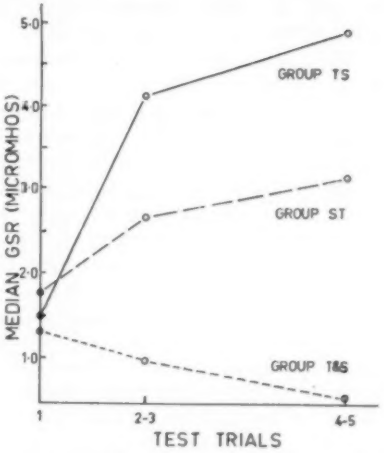

FIG. 1. Performance curves under forward conditioning (Group TS), backward conditioning (Group ST), and pseudoconditioning (Group T & S).

was retained. The CS-alone trials (T) used as test trials in Group T & S were those which were preceded, on the average, by 6, 9, 13, and 15 shocks, as in the experimental groups. In addition, the whole series was begun with a CS-alone trial.

RESULTS

The procedure adopted with Groups TS and ST did not allow any measure to be taken of the GSR to the tone before any UCS-alone trials or CS-UCS trials. A number of experiments conducted in the University of Sydney laboratory have shown that the size of a GSR to a tone varies with the duration and intensity of the stimulus but that the median GSR before any training is approximately 1.5 micromhos for a 20-msec., 60-db. tone; a value of this order was therefore derived for each *S* in the two experimental groups by taking 15% of the mean score over the last three preliminary shock trials. The validity of this procedure is confirmed

by the measured median value of 1.3 micromhos in Group T & S.

The performance curves obtained in the experiment are shown in Fig. 1 and it may be observed that the rise in performance for the forward and backward groups (Groups TS and ST) is accompanied by a slight fall in the performance of the pseudoconditioning group (Group T & S). It is emphasized that the number of shocks administered before each block of test trials is constant over the three groups; in average terms, and excluding the preliminary shock-alone trials, there were 2.5 shocks before the first block of test trials (Trials 2–3 in Fig. 1), after two training trials in Groups TS and ST, and nine shocks before the second block of test trials (Trials 4–5 in Fig. 1), after six training trials in Groups TS and ST.

Because the scores were not normally distributed, nonparametric tests were used. An H test (Edwards, 1954) was applied to the total GSR for the first and second blocks of test trials combined and it showed that the three groups differed significantly ($H = 27.79$, $P < .01$). This was followed by separate U tests (McNemar, 1955) comparing Groups TS and ST ($z = 1.39$, $P = .16$) and Groups ST and T & S ($z = 3.96$, $P < .001$). The trend in the three groups combined was examined over test trials by means of Mood's test (McNemar, 1955) and it proved significant ($\chi^2 = 7.30$, $P < .05$). Separate Mood tests then showed significant upward trends in Groups TS and ST ($\chi^2 = 10.90$ and 12.10, respectively, with $P < .01$) but no trend in Group T & S ($\chi^2 = 5.20$, $P > .05$). One other point of interest is that the difference between Groups TS and ST was significant on Test Trial 2 according to a U test ($z = 2.06$, $P = .04$). This is the stage at which the greatest rise in performance typically occurs in GSR conditioning (Champion, 1961; Moeller, 1954), as with many other responses.

Discussion

The results point to a difference between backward conditioning and pseudoconditioning of the GSR as defined by the procedures used with Groups ST and T & S, respectively. The performance of Group ST showed a rise whereas that of Group T & S showed a nonsignificant fall. The initial improvement of the Group ST may not have been as great as that of the Group TS, but the two groups did not differ significantly over training as a whole. The general conclusion to be drawn is that the pseudoconditioning of the GSR is a basically different process from that occurring in forward and backward conditioning. A limitation to be placed on this conclusion at present is that the forward or backward interval between CS onset and UCS onset should not exceed 750 msec. In a pseudoconditioning control group Moeller (1954) used a backward conditioning procedure with the UCS-CS interval varying between 3000 and 5000 msec. He obtained no rise in performance over 23 trials on Day 2 of the experiment but did not report the effect of the 15 trials which all *Ss* received on Day 1.

The present data lend weight to the possibility that UCS-CS trials may produce some degree of true conditioning in the case of a long-latency response such as the GSR. This outcome is in direct opposition to the prediction of Hilgard (1951) that a longer forward interval would be more appropriate with a long-latency response, which has been disproved (Moeller, 1954; White & Schlosberg, 1952). Further consideration of the initial argument on the definition of contiguity shows that a backward gradient of reinforcement must be assumed to operate in the conditioning of the GSR no matter what aspect of the UCS is taken as reinforcement unless the offset of the UCS is delayed an unusually long time. This is generally true in

backward conditioning and applies in forward conditioning as long as the latency of the CR exceeds the interval between CS onset and UCS offset. If the concept of "backward reinforcement" is unacceptable then it presumably must be postulated that the conditioning of the GSR occurs without reinforcement. A possible reason for the failure to obtain true conditioning of a short-latency response with UCS-CS trials may be found if the appropriate latencies of the responses are considered. If the latency of the response is less than 500 msec., as with hand withdrawal or the eyeblink, then forward gradients of contiguity and reinforcement may be assumed in forward conditioning but both gradients must be taken as operating backwards in backward conditioning, again provided that UCS offset is not unduly delayed.

Summary

A comparison was made of forward, backward, and pseudoconditioning of the GSR in the expectation that a backward procedure would produce true conditioning in such a long-latency response. The expectation was confirmed, for whereas the performance of the first two groups increased with paired CS-UCS or UCS-CS trials, that of the pseudoconditioning group decreased to a nonsignificant degree with comparable numbers of unpaired CS and UCS trials. Reasons were advanced for the failure on the part of other workers to obtain conditioning of a short-latency response with UCS-CS trials.

References

Champion, R. A. Supplementary report: Interpolated UCS trials in GSR conditioning. *J. exp. Psychol.*, 1961, in press.

Edwards, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

McNemar, Q. *Psychological statistics.* (2nd ed.) New York: Wiley, 1955.

Mason, J. E. The joint action of contiguity and reinforcement in classical conditioning. Unpublished doctoral dissertation, University of Sydney, 1959.

Moeller, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48,** 162–166.

Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

Spooner, A., & Kellogg, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, **60,** 321–334.

Underwood, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.

White, C. T., & Schlosberg, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43,** 357–362.

(Received May 6, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 62–69

ACQUISITION OF PERCEPTUAL RESPONSES AS A FUNCTION OF LOADING, LOCATION, AND REPETITION [1]

HARRY W. KARN AND LEE W. GREGG

Carnegie Institute of Technology

The purpose of the present investigation was to determine the ways in which the visual perception of elements of a stimulus complex is influenced by the amount of experience with the complex and by specific features of the complex designated as location and loading.

The critical aspect of the experiment was the simultaneous presentation of three circles under spatial and temporal conditions which made the discrimination of a distinguishing binary cue (the presence or absence of a dot within each circle) difficult or impossible for all circles of the set. If the duration of the stimulus was made long enough, *S* could, by scanning the three circles, report accurately upon the nature of the individual circles. Given such an opportunity, *S* could learn that for one of the circles the value of the binary cue varied, while the values for the other circles remained constant over a series of presentations. These features of the stimulus field, if learned, could make it possible to report accurately at stimulus durations that precluded the possibility of scanning the entire stimulus field.

It was supposed that through such an experimental plan, evidence would be obtained to determine the ways perceptual responses are acquired by providing precise control over the nature of learnable cues and conditions of detection.

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-770.

METHOD

Subjects.—The *S*s were 168 male college students, six in each of 24 experimental groups and 12 in each of two control groups.

Apparatus and procedure.—Three circles located at the points of an imaginary equilateral triangle were simultaneously presented to *S* on a screen 4 × 5 ft. One side of the triangle was 30 in. in length, subtending at *S*'s viewing distance approximately 15° of visual angle. The outside diameter of each circle was 1 in. corresponding to a visual angle of 2′ of arc. The discrimination required of *S* was the presence or absence, within each circle, of a dot subtending approximately 1′ of arc.

For the experimental *S*s all possible combinations of loading and location were used. Location refers to the position of the circle at the apex of the triangle (Location A), at one end of the base to *S*'s right (Location B), and at the other end of the base to *S*'s left (Location C). Loading refers to complete presence (positive loading), complete absence (negative loading), or random occurrence (random loading) of the dot within the circle at each location. The nature of the six configurations (Cs) presented to experimental *S*s are as follows: C_1 = A+, B−, Cr; C_2 = A+, Br, C−; C_3 = A−, Br, C+; C_4 = A−, B+, Cr; C_5 = Ar, B+, C−; C_6 = Ar, B−, C+.

Control *S*s observed circles in which the presence or absence of a dot was randomly determined for each of the three circles.

For each of the six Cs, four groups of experimental *S*s received 30, 20, 10, or 0 prepresentations, respectively, of the C at 1-sec. exposure followed by 60 presentations at .1-sec. exposure.

All *S*s were seated in an armchair to which was attached a panel with three toggle switches. Each switch had a neutral center position and could be thrown forward to indicate the presence of a dot or back toward *S* to indicate the absence of a dot. Small ink drawings of circles with or without the

dot were placed in the appropriate directions for each switch. The switches were mounted in a triangular array to provide maximum S-R compatibility.

The experimental room was dimly illuminated at all times by a single lamp located above *S*. The screen brightness was approximately 5 ml.

The instructions, which were read by *E*, informed *S* that three circles at the points of an imaginary equilateral triangle would be flashed on the screen before him and that sometimes one or more of the circles would contain a dot. A report on all circles was required with instructions to guess if unsure. No time limit for the response was imposed. After recording *S*'s responses by noting the order in which indicator lights came on, *E* asked *S* to return the switches to the neutral positions and to get ready for the next presentation.

Results

The number of errors for each target for the prepresentations at 1 sec. was obtained for each *S*. Group mean errors at each location for the 36 *S*s of the experimental groups receiving 10, 20, or 30 prepresentations and for the 12 control *S*s who received 30 prepresentations are given in Table 1. The same error data for the experimental *S*s broken down according to the loading characteristics of the circles are also presented.

No errors were observed beyond the tenth presentation for any of the groups receiving preliminary experience at 1 sec. The mean total errors shown in Table 1 could therefore be considered as data from the first block

TABLE 1

Group Mean Errors during Preliminary Presentations at 1 Sec.

Group	n	Location			Loading			Mean Total Errors
		A	B	C	+	−	r	
Exp. 10	36	1.3	1.1	1.1	1.2	1.2	1.1	3.5
Exp. 20	36	1.0	1.0	1.0	1.0	1.0	1.0	3.0
Exp. 30	36	1.1	1.2	.9	1.2	1.0	1.0	3.2
Cont. 30	12	1.2	1.1	1.0	random			3.3

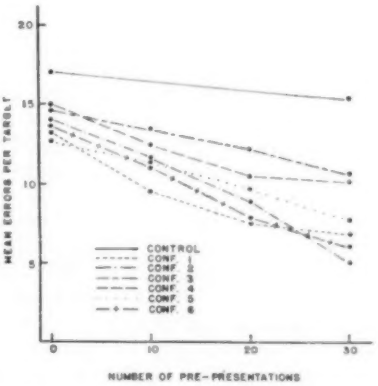

Fig. 1. The effect of preliminary experience on perceptual performance for six "loaded" configurations by experimental *S*s and an "unloaded" configuration by control *S*s.

of 10 prepresentations with 30 opportunities for error. It is clear from inspection of the table that regardless of location, loading, or the total number of prepresentations that *S*s finally received, no systematic variation in errors exists. The magnitude of the errors was uniformly low, about 10%, in every case.

For the 60 presentations at .1 sec. the number of errors for each target was obtained for each *S*. Group mean errors per target are shown in Fig. 1. The individual curves are distinguished by the particular combinations of loading at the three locations. These combinations were referred to earlier as Configuration (C). The solid curve of Fig. 1 connects the means of the two control groups.

It is apparent that preliminary practice had little effect on the performance of control *S*s. The mean for the Control 0 group is 16.92 and that for the Control 30 group is 15.11, a difference of 1.81 errors. An analysis of variance for the control group data indicated that the reduction in errors as a function of number

of prepresentations at .1 sec. was not significant ($F = 1.16$, $df = 1/44$; $.25 < P < .50$). The analysis of these data also showed that the effect of location was not significant ($F = 2.13$, $df = 2/44$; $.10 < P < .25$), nor was the interaction significant ($F = .17$, $df = 2/44$; $P = <.50$).

The control curve is based on two points at the extremes of the range of prepresentations, 0 and 30. Instead of running 6 *Ss* as in the experimental groups, the number of *Ss* was doubled for the extreme points to increase the power of the statistical test. Based on the *Ss* within error variance, the number of *Ss* in each group required to detect a difference between the Cont. 0 and Cont. 30 means of 1.81 is 145.

For the *Ss* of the experimental groups, there is a decrease in errors as a function of the variable, number of prepresentations. Separate analyses of variance were performed for the data of the experimental groups that had received the same C. Except for C_3, Number of prepresentations was a significant source of variation. The *F*s for this source in the analyses were 4.78, 8.27, 1.23, 5.67, 3.58, and 4.73, with 3 and 20 *df* for C_1 through C_6, respectively. The error variances for these analyses were homogeneous. The best estimate of the population variance for these treatments (36.52) was significantly smaller than that found in the control group analysis (58.80).

Loading and Location are necessarily confounded within any given configuration. The main effects of these variables can be demonstrated, however, by summing over Cs. The nature of these effects can be seen in the curves of Fig. 2. The left hand panel of this figure presents group mean errors per target as a function

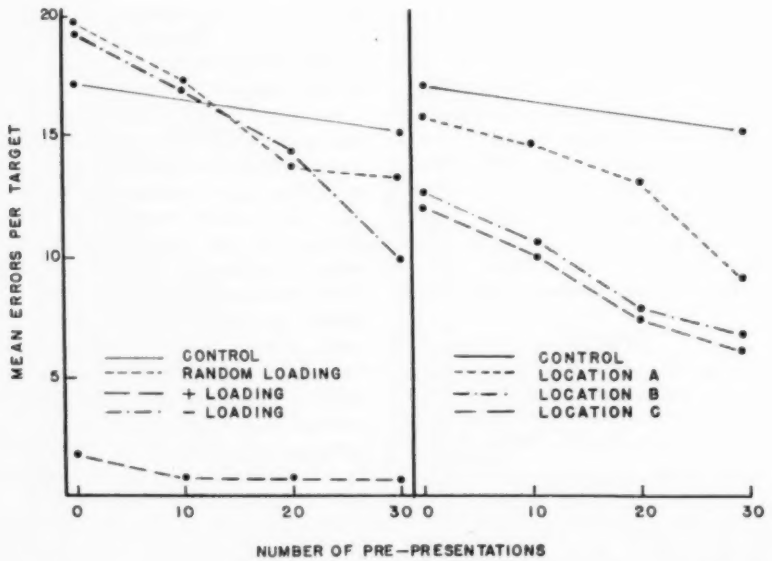

FIG. 2. Changes in perceptual performance as a function of preliminary experience with loading as the parameter (left panel) and with location as the parameter (right panel).

of the number of prepresentations at 1 sec. with Loading (regardless of Location) as the parameter. In the right hand panel of Fig. 2 the same data are used to compute group means with Location (regardless of Loading) as the parameter for the curves. In this figure the solid curves are based on the same means as the control curves of Fig. 1.

In order to assess the significance of the main effects of loading for the experimental groups, the data were treated as follows. Two latin squares were constructed so that the rows of the squares exhausted the set of six Cs. The first square contained the data for those experimental groups that received C_1, C_3, and C_5; the second square was for C_2, C_4, and C_6. For each square then, the columns represented Target Location; the rows represented Configurations as previously defined; and the letters represented Loading. It should be emphasized that although these squares were constructed, rather than randomly selected, together they provide the same information as the complete set of 12 3×3 latin squares. This is true because independent groups of *Ss* were randomly assigned to the rows. Hence, row permutations of any kind cannot influence or bias the variance estimate and all possible column permutations are contained in the six Cs. The variable Number, refers to the prepresentations at 1 sec. and provides an additional source of variation orthogonal to the latin square sources. The summaries for the separate analyses are presented in Table 2. This form for presenting the outcome of the statistical treatment was selected in preference to a single combined analysis because it provides a comparison of what could be considered two independent sets of observations, each based on the performance of 72 *Ss*. It should be noted that all basic sources of error are of the same magnitude. For each analysis the insignificant $Error_1$ (*Ss* within groups) was pooled with the residual error, $Error_2$, to yield a single error term appropriate for all subsequent *F* tests.

TABLE 2

ANALYSES OF VARIANCE OF ERRORS PER TARGET

Source	*df*	Configurations 1, 3, 5 Square 1		Configurations 2, 4, 6 Square 2	
		MS	*F*	*MS*	*F*
Loading (L)	2	5480	156.57*	4669	155.63*
Target location (T)	2	467	13.34*	478	15.33*
Configurations (C)	2	221	6.31*	123	4.10*
Square uniqueness	2	297	8.49*	699	23.30*
Number (N)	3	329	9.40*	547	18.23*
N × L	6	88	2.51*	106	3.53*
N × T	6	48	1.37	21	.70
N × C	6	17	.49	21	.70
N × U	6	25	.71	13	.43
$Error_1$ (*Ss* within)	60	40	—	33	—
$Error_2$ (N × *Ss* within)	120	32	—	28	—
Pooled error sources	180	35	—	30	—
Total	215				

* $P < .05$.

The same inferences would follow from consideration of either of the analyses. The main effects of Loading, Target Location, Configurations, and Number of Prepresentations were significant beyond the .05 level in both.

Square uniqueness and the N × L interaction were also significant. Figure 2 is based on the combined data for all six Cs but nonetheless reflects accurately the origin of the N × L interaction. The curve for positive loading (dot always present) as a a function of number of prepresentations is initially low and remains so. The group means are less than two errors per target for all levels of preliminary experience. The curves for the conditions of random and negative loading show a sharp and reasonably linear decrease. These

TABLE 3

MEAN ERRORS PER TARGET FOR EACH CONFIGURATION, LOADING, AND LOCATION

C	(+)	(−)	(r)	Mean
1	.7 (A)	11.4 (B)	15.7 (C)	9.3
3	.3 (C)	22.4 (A)	15.2 (B)	12.3
5	1.1 (B)	12.1 (C)	17.9 (A)	10.3
Mean	.7	15.3	16.3	10.6
2	.7 (A)	11.1 (C)	18.0 (B)	9.9
4	1.8 (B)	22.7 (A)	12.3 (C)	13.3
6	1.5 (C)	11.4 (B)	16.5 (A)	9.8
Mean	1.3	15.0	15.6	10.7

Note.—Each entry in Table 3 is the mean of 24 *Ss*, six from the four experimental groups within each C. Using the pooled error estimates from both of the analyses of Table 2 gives a combined value of 32.5 as the σ_m per experimental unit. The $\sigma_{diff.}$ in the main body of the table is given by the expression $\sqrt{2}ms/r$. The computed value is 1.65. A difference as large or larger than 3.41 is significant at the .05 level.

marked differences in slope give rise to the significant interaction term.

The interpretation of the source called Square Uniqueness requires special comment. In the usual analysis of an unreplicated 3×3 latin square, the 8 *df* would be partitioned into 2 *df* each for row, column, and letter effects with 2 *df* remaining that may be used as an error term if there are no intrinsic interactions. Replicating a single square, as is the case for each of the squares in the present study, makes it possible to determine whether or not the particular square selected contains a unique pattern of confounded interaction effects if such interaction exists. In squares as small as the 3×3 one cannot suppose that confounded interaction effects would tend to cancel as they do in the larger squares. The significant Square Uniqueness effects in the present squares, therefore, means that there is some form of interaction among the variables Loading, Target Locations, and Configurations.

Table 3, which presents the means obtained by summing over *Ss* of all four experimental groups receiving a particular C, makes it possible to account for the significant effect that Square Uniqueness indicates is present. In general, positive loading yields the fewest errors followed by the condition of negative loading with the random loading condition producing the greatest number of errors. This pattern obtains for C_1, C_2, C_5, and C_6. C_3 and C_4 depart radically from this pattern, reversing the error magnitudes associated with negative and random loadings. Moreover, the means for the negative loadings are the largest in the entire table. C_3 and C_4 have one thing in common—the negatively loaded circles are at Location A, the apex of the triangle.

The preceding analyses indicate that without specific instructions to do so *Ss* learn to utilize the cues inherent in the loaded fields to improve the accuracy of their perceptual reports. This learning is a function of the amount of prior experience with the target field.

Does proficiency of performance change during the presentations at the .1-sec. exposure even without prior experience? Are the rate and extent of such changes dependent upon amount of prior experience?

To answer these questions blocks of 20 consecutive presentations within the 60 presentations at .1 sec. provided three error measures for each *S*. These measures represent total errors for all targets and can take on values from 0 to 60 errors. Means based on 36 *Ss* in the experimental groups and 12 *Ss* in the control groups are plotted in Fig. 3 as a function of the ordered blocks of presentations. Number of prepresentations at 1 sec. is the parameter.

Summary tables for the separate

analyses of variance for the control and experimental groups are presented in Table 4. The main effects of Number of prepresentations and Configurations are significant in the case of the experimental groups. Number is not significant as a source of variation in the control group analysis. These main effects have their counterparts in the latin square analyses and control group analysis mentioned earlier.

The main effect of Blocks is of primary concern in the present analyses. For the experimental groups, the Block effect and the B × N interaction are significant. Error reduction occurs, and does so at a differential rate depending upon the number of prepresentations, within the 60 presentations at .1 sec. For the control groups, neither the main effect of Blocks nor the B × N interaction was significant at the .05 level.

When the number of prepresentations at 1 sec. is 0, the sharpest drop in errors occurs over the blocks of presentations at .1 sec. With increasing numbers of prepresentations, the rate of error reduction within the 60 presentations at .1 sec. decreases as indicated by the slopes of the curves for the experimental groups in Fig. 3.

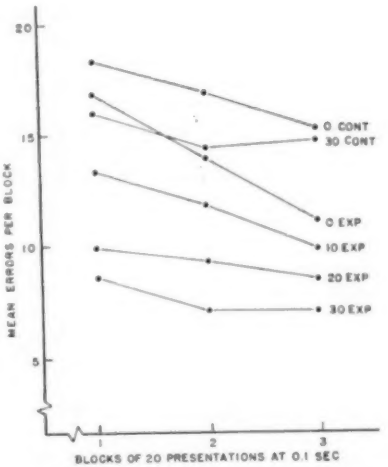

FIG 3. Changes in perceptual performance within 60 trials at .1 sec. for control and experimental *Ss* receiving differing amounts of preliminary experience.

DISCUSSION

Under the conditions of the present experiment, a reduction in errors in reporting the state of three simultaneously presented targets occurred. The magnitude of the error reduction was a function of amount of prior experience

TABLE 4

ANALYSIS OF VARIANCE OF ERRORS PER BLOCK

Source	Experimental Groups			Control Groups		
	df	*MS*	*F*	*df*	*MS*	*F*
Number (N)	3	863.42	23.68*	1	58.68	.99
Configurations (C)	5	136.56	3.75*			
N × C	15	18.40	.50			
$Error_1$ (*Ss* within)	120	36.46	4.52*	22	58.80	6.30*
Blocks (B)	2	291.88	36.21*	2	28.35	3.04
B × N	6	47.32	5.87*	2	7.68	.8
B × C	10	6.34	.79			
B × N × C	30	3.63	.45			
$Error_2$	240	8.06	—	44	9.33	—
Total	431			71		

* $P < .05$.

with the targets. This kind of behavioral change is considered to be "perceptual learning" because the reduction in errors depended primarily on stable, and hence learnable, properties associated with the stimulus complex. Even in the absence of preliminary observation where conditions were favorable to seeing the relationships among the stimulus elements, error reduction attributable to the stable cues took place.

In what appears to be a relatively simple observational task some rather complex phenomena emerge. The initial supposition on which the experimental plan was based was that *S* would notice the stable nature of the positively and negatively loaded targets and then select for his initial fixation point the area in which the randomly varying target would appear. In this way nearly perfect performance could be expected. Moreover, having seen this solution, *S*'s response patterns should show the sudden, insightful shift typically associated with problem solving tasks. Such was not the case. Instead, the improvement, in general, was slow and incremental for the negatively loaded and randomly varying targets. The only marked discontinuity was the extremely rapid reduction in errors for the positively loaded target. This superior performance with respect to the positively loaded target coupled with the relatively high number of errors for the negatively loaded target, especially when the latter occupied a nonpreferred position in the stimulus field, poses a paradox since both elements provide the same degree of stability as a cue for learning.

It is evident that a complete explanation of the observed relationships will depend upon processes other than those implied in the original formulation of the task. Certainly, *S*s will and do acquire information about the relationships that exist within the stimulus complex. This information provides the background against which the perceptual responses to the complex are made. However, even in an arrangement as limited as the three binary-valued elements used in this study, several additional cognitive and response processes must play a role.

First, consider the habitual ways of scanning the stimulus field (Carmichael & Dearborn, 1947). When more than one element is simultaneously presented, attention will be divided, shifting element to element if time allows. Depending on the internal relationships within the stimulus complex, a variety of possible perceived relationships could be educed as a function of the relative positions of the elements. Or what of the effect of set? Perhaps a partial description of the perceptual learning phenomena reported here rests in the establishment of perceptual sets akin to the cognitive sets in certain problem solving experiments (Gardner & Runquist, 1958; Kendler, Greenberg, & Richman, 1952). Of a similar nature is the finding that response sets can be established which operate in the absence of any physical stimulus (Goldiamond & Hawkins, 1958). Not unlike the response sets are the factors related to *S*'s criteria for reporting and these, as has been suggested, may interact with the more basic variables such as intensity which determine the probability of seeing (Tanner & Swets, 1954).

At present, full-scale investigations of various instructional sets and of the "probability of seeing" effects on the kind of perceptual learning reported here are being conducted. Until such a time as perceptual theory can bring together in coherent sequence the complex processes that generate such behaviors, the present data should stand as a caution to investigators who elect to present tachistoscopically complex figures, words, and the like and from such presentations draw inferences concerning their perceptual consequences.

Summary

The aim of this study was to determine the ways perceptual responses are acquired by providing precise control over the nature of learnable cues and conditions of detection.

Three circles located at the points of an imaginary equilateral triangle were simultaneously presented to *S*s on a screen by means of a tachistoscope with instructions

to report presence or absence of a dot within each circle. Loading refers to complete presence, complete absence, or random occurrence of the dot at a location.

For each of six possible combinations of loading and location, four groups of experimental *Ss* received 60 presentations at .1-sec. exposure following either 0, 10, 20, or 30 preliminary presentations at 1 sec. Control *Ss* observed circles in which the presence or absence of a dot in all circles was randomly determined. Since conditions were such that it was impossible to observe all circles at .1 sec., changes in performance may be attributed to experience at the 1 sec. exposure.

There was a significant reduction in errors as a function of the number of preliminary 1-sec. presentations under all conditions of loading and location. But both location and loading influenced the magnitude of the error, and there was a significant interaction between loading and location.

Tentative explanations of the obtained differential loading and location effects include habitual scanning procedures, perceptual sets, and subjective criteria used in reporting.

References

CARMICHAEL, L., & DEARBORN, W. H. *Reading and visual fatigue.* Boston: Houghton Mifflin, 1947.

GARDNER, R. A., & RUNQUIST, W. N. Acquisition and extinction of problem-solving set. *J. exp. Psychol.*, 1958, **55**, 274–277.

GOLDIAMOND, I., & HAWKINS, W. F. Vexier-versuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, **56**, 457–463.

KENDLER, H. H., GREENBERG, A., & RICHMAN, H. The influence of massed and distributed practice on the development of mental set. *J. exp. Psychol.*, 1952, **43**, 21–25.

TANNER, W. P., & SWETS, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

(Received May 10, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 70-79

ASSOCIATIVE FACTORS IN VERBAL TRANSFER [1]

JARVIS BASTIAN [2]

University of Minnesota

In the two decades since the publication of Gibson's (1940) well-known analysis of verbal transfer in terms of generalization and differentiation, an impressive amount of research has been directed towards tracing out the effects of verbal generalization in a wide range of transfer situations. These efforts have produced a body of results of sufficient breadth and uniformity to permit the construction of fairly general statements of the experimental effects of verbal generalization, such as those provided by Osgood (1953) and Underwood (1949).

However, comparatively little has been done to determine the underlying processes responsible for these effects. It is generally conceded that verbal generalization, aside from that which is ascribable to the phonetic or orthographic properties of the experimental materials, is the result of processes quite different from those governing primary generalization. Nevertheless, the feeling persists that there must be parallels between the two phenomena, particularly in regard to the roles played by similarities in the verbal domain and the physically specifiable similarities associated with primary generalization.

It is perhaps for this reason that the notion of "meaning" is so often invoked in theoretical discussions of verbal generalization, because "similarities in meaning" apparently provide a basis for verbal generalization that is analogous to the similarities in physical attributes involved in primary generalization.

In spite of the extremely formidable logical and linguistic problems that are involved in the notion of "similarities of meaning," its use in this context has been strengthened by a considerable amount of experimental evidence clearly relating the degree to which verbal materials are judged to be similar in meaning to the degree of generalization obtaining between these materials. Haagen (1943), Morgan and Underwood (1950), and Underwood (1951) have all reported transfer results in paired associate training which indicate that generalization effects increase as a function of the degree to which words in comparable interlist positions are judged to be similar in meaning.

However, the results of these studies are possibly contaminated by another, uncontrolled, factor in the experiments. The experimental variable, degree of similarity in meaning, was manipulated in all of these experiments by selecting words from a set of 400 pairs of adjectives for which Haagen (1949) had collected ratings of degree of synonymity. But Haagen also obtained, from a separate group of judges, ratings of the degree of associative connection between these same word pairs. The correlation between the scale values derived from the two sets of ratings was .90, which was only slightly less than their individual reliability coefficients. Thus, in the experiments cited, the generalization results can be attributed as much to the associative factor as to the similarities in meaning, if, indeed, these appear as separate factors in these studies.

Although degree of semantic similarity and degree of associative connection are

[1] This report is based on part of a PhD dissertation presented to the Graduate School of the University of Minnesota. The author is deeply indebted to W. A. Russell and J. J. Jenkins for their advice and encouragement.

[2] Now at Haskins Laboratories, New York, New York.

evidently highly correlated, it might be possible to manipulate them separately because of the considerably broader range of interverbal relations covered by associative tendencies. The exploration of this possibility is highly desirable, for McClelland and Heath (1943) and Russell and Storms (1955) have reported transfer experiments which suggest that associative factors may be fundamentally involved in the production of verbal generalization. These studies both demonstrated pronounced generalization effects across paired associate lists containing words in analogous interlist positions which were linked by associative tendencies, as inferred from word association norms, but which were not semantically related in any acceptable sense.

The present experiment was designed to assess the separate and combined generalization effects produced by the two different interverbal relations in order to provide some further information that might lead to a clearer conceptualization of the factors mediating verbal generalization.

Method

The strategy of this research required a situation that provided a sensitive index of verbal generalization, and one in which the factors of associative strength and similarity of meaning could be separated and their effects compared. For this purpose a transfer of training situation was selected that involved the successive learning of two lists of paired associates containing identical stimulus elements. The basis for the selection of this experimental situation is that in the successive learning of paired-associate lists, each containing the same set of stimulus items, but different response items, generalization between the responses results in positive transfer, whereas the lack of generalization results in negative transfer (cf. Osgood, 1953; Underwood, 1949). In addition, parallel experiments conducted by Haagen (1943) have indicated that this situation is more sensitive to generalization effects than those in which responses are held constant across lists and stimuli varied, or those in which both stimulus and response elements are varied.

Materials.—With these considerations in mind, lists of 12 paired associates were constructed with the same set of words appearing as stimulus elements in each. The response elements in List 1 were taken from the stimulus words used in the Kent-Rosanoff free association test.

Each response element in the other lists was related to a given response element in List 1 by one of the following three relationships: Type A elements were highly frequent word associates of the List 1 response words, but were judged to bear little or no similarity to them in meaning; Type S elements were very infrequent word associates of the List 1 response words, but were judged to be highly similar to them in meaning; Type C elements were control words which were neither associates of, nor judged to be similar in meaning to the List 1 response words.

The measure of frequency of association between the List 1 response words and the response words in the other lists was obtained from the frequency with which these words appeared as responses to the List 1 words in the Kent-Rosanoff norms provided by Russell and Jenkins (1954). The measure of similarity of meaning between the List 1 response words and those in the other lists was obtained from median judgments provided by a group of 32 judges not otherwise pariticipating in the experiment. The judges were asked to rate each pair of words on a four-point scale ranging from a minimum at "0" (defined for them as: "no similarity or even opposite in meaning") to a maximum at "3" (defined as: "identical or very similar in meaning").

From a much larger pool of word pairs for which these measurements were available, 36 words were selected which most closely matched the properties of each of the Type A, S, or C relationships to the List 1 response words. (It turned out to be exceptionally difficult to discover words which met these combined requirements. Evidently, if a pair of words in general use is judged to be similar in meaning, it is also quite likely that each will be evoked by the other in a word association test.) Lists 2, 3, and 4 were then constructed by pairing each of these words with the same stimulus word that had been paired with their counterparts in List 1, so that Lists 2, 3, and 4 each contained four response words representing each of the three types of interlist relationships. The composition of the lists and the type of relation each of the response words in Lists 2, 3, and 4 bore to the List 1 response words are presented in Table 1.

TABLE 1

COMPOSITION OF LISTS AND INTERLIST RELATIONS

Stimulus Words All Lists	List 1 Responses	List 2		List 3		List 4	
		Responses	Type	Responses	Type	Responses	Type
BREAD	HIGH	LOFTY	S	LEFT	C	LOW	A
COLD	STEM	TRUNK	S	FLOWER	A	PROVE	C
EAGLE	SICKNESS	HEALTH	A	BROOK	C	DISEASE	S
FRUIT	QUIET	LOUD	A	HAND	C	SILENCE	S
HAMMER	SOUR	SWEET	A	ACID	S	NARROW	C
JUSTICE	DREAM	IMAGINE	S	MARKET	C	SLEEP	A
MUTTON	LONG	SAND	C	SHORT	A	TALL	S
NEEDLE	KING	QUEEN	A	CHIEF	S	CLOSE	C
OCEAN	BOY	NORTH	C	LAD	S	GIRL	A
SLOW	LIGHT	MINE	C	DARK	A	GAY	S
TABLE	MAN	CONTAIN	C	HUMAN	S	WOMAN	A
THIRSTY	HARD	DIFFICULT	S	SOFT	A	LITTLE	C

All words were carefully screened to eliminate any other associative or semantic relations within or between the lists, and to eliminate orthographic similarities and words having low frequencies of usage.

In order that differences in the extent of generalization might be confidently inferred from differences in performance in the transfer task, it would first be necessary to establish that such differences in performance could not be attributed to intrinsic differences in the ease with which different types of pairs are learned, irrespective of the effects of the prior learning of the related pairs. For this reason it was desirable to have different groups of *Ss* learn the lists in two different sequences (Sequences I and II) so that the subsequent task in one sequence would be the same as the prior task in the other sequence. It was also thought desirable to investigate the generalization effects produced by the various interlist relations in the recall of the original pairs. Therefore, the order of tasks presented in Sequence I was: learn List 1; learn List 2, 3, or 4; recall List 1. The comparable order in Sequence II was: learn List 2, 3, or 4; learn List 1; recall List 2, 3, or 4.

However, it should be appreciated that the associative relation is not necessarily sym-

TABLE 2

PERCENTAGE ASSOCIATION AND MEDIAN RATINGS OF SIMILARITY TO LIST 1 RESPONSE WORDS FOR WORDS IN EACH TYPE OF INTERLIST RELATION

Type A	Association		Similarity	Type S	Association		Similarity	Type C	Association		Similarity
	Seq. I	Seq. II			Seq. I	Seq. II			Seq. I	Seq. II	
DARK	64.2	82.2	0.53	ACID	0.2	6.5	1.92	BROOK	0.0	0.0	0.50
FLOWER	39.9	1.0	0.84	CHIEF	0.1	0.0	2.38	CLOSE	0.0	0.0	0.52
GIRL	76.2	69.8	0.57	DIFFICULT	0.5	65.5	3.11	CONTAIN	0.0	0.0	0.53
HEALTH	37.3	24.8	0.53	DISEASE	0.1	19.4	2.44	HAND	0.0	0.0	0.50
LOUD	34.5	6.7	0.52	GAY	0.0	0.0	2.07	LEFT	0.0	0.0	0.50
LOW	67.0	64.0	0.53	HUMAN	0.3	36.0	1.65	LITTLE	0.0	0.0	0.50
QUEEN	74.5	71.0	0.99	IMAGINE	0.0	22.3	2.53	MARKET	0.0	0.0	0.53
SHORT	75.2	33.3	0.52	LAD	0.5	69.1	3.16	MINE	0.0	0.0	0.50
SLEEP	44.9	4.7	0.94	LOFTY	0.1	64.0	2.99	NARROW	0.0	0.0	0.52
SOFT	66.9	44.1	0.52	SILENCE	0.6	53.2	3.27	NORTH	0.0	0.0	0.50
SWEET	56.3	43.1	0.53	TALL	0.5	0.7	1.82	PROVE	0.0	0.0	0.53
WOMAN	76.1	61.1	0.57	TRUNK	0.7	0.0	2.25	SAND	0.0	0.0	0.50
Mean	59.4	42.8	0.63		0.4	28.1	2.47		0.0	0.0	0.51

metric. That is, the extent to which Word A evokes Word B as an association response is not necessarily the same as the extent to which Word B evokes Word A. Thus it was necessary to determine the frequency with which the response words in List 1 tended to be given as association responses to their counterparts in List 2, 3, or 4. This frequency was inferred from additional association norms collected by Russell and Jenkins (1956) and from norms collected by *E* from 139 *Ss* not otherwise participating in the experiment.

Table 2 presents the interlist association and similarity properties in both sequences for words in each type of interlist relation. It can be seen that on the average the properties of the interlist relations in the converse order (Sequence II) were different from those obtaining in Sequence I in that the response words participating in the Type S relation were characterized by a substantial degree of interlist associative linkages, whereas the associative linkages between the words of Type A were somewhat weaker than those in Sequence II. Since of necessity, the similarity in meaning relation is symmetric, the words representing the Type S interlist relation in Sequence II combined to a considerable extent the properties of both Type A and Type S relations.

Procedure.—Seventy-two *Ss*, students in an introductory psychology course, participated in the experiment, 36 learning the lists in each sequence. In order to obtain a measure of the effect of the different lists, *Ss* were further divided into three equal groups within each sequence, according to the particular lists they were to learn during the different phases of the experiment. In Sequence I *Ss* in Groups 1, 2, and 3 learned Lists 2, 3, and 4, respectively, as their second lists, and in Sequence II Lists 2, 3, and 4 were learned as their original lists by Groups 1, 2, and 3, respectively.

After a short period of pretraining on a practice list of nonsense syllable-digit paired associates, *Ss* proceeded to learn two of the experimental lists by the anticipation method, with instructions to pronounce aloud both stimulus and response words on all occasions. The materials were presented on a Hull-type memory drum arranged so that the stimulus words were exposed for 2 sec. and then were joined by the response words for another 2 sec. Two seconds intervened between the presentation of each pair, and 8 sec. separated successive trials. The 12 pairs in each list were presented in five different orders.

The *Ss* were instructed to correct all mistakes aloud and to continue practice on the lists until all response words had been correctly anticipated in a single trial. When this criterion had been reached on the original list, a shutter arrangement on the apparatus was shifted to expose the second list, and after a brief period for instructions lasting approximately 1 min., practice began immediately on the second list and continued until the same criterion was met. Then, immediately following an instructional period of the same duration in which *Ss* were told to recall as many of the original responses as possible on the first trial, the shutter arrangement was shifted back to the original list, and the recall test began and continued until all responses had been correctly anticipated in one trial.

Results

Original list learning.—The number of trials required to reach the criterion of correctly anticipating all responses in the original lists was analyzed for any overall differences in the ease with which the lists were learned, irrespective of the types of interlist relations in which their component responses would subsequently participate. The mean numbers of trials needed to attain this criterion for Groups 1, 2, and 3 in learning List 1 in Sequence I were 13.17, 11.83, and 12.67, respectively. The groups learning Lists 2, 3, and 4 in Sequence II required 14.38, 11.92, and 11.00, respectively. The differences among the group performances on these lists cannot be considered reliable ($F = .92$, $df = 2/66$), nor were the other sources of variation significant.

The total numbers of correct anticipations given by all *Ss* in learning the original lists in each sequence were analyzed for differences among the different types of interlist relations. Since each list contained words from each category, measurements of *Ss'* performance on them cannot be considered independent. Therefore, Collier's (1958) modification of the model for the analysis of observations

correlated on one dimension presented by Kogan (1948) was applied.

The mean total correct anticipations of response words of Types A, S, and C, respectively, were 31.69, 31.89, and 32.39 in Sequence I, and 32.00, 31.91, and 31.08 in Sequence II, and 31.84, 31.90, and 31.73 for the two sequences combined. The differences among these means cannot be considered reliable, nor were any reliable differences found between groups or sequences (all Fs $< .80$).

However, significant Type×Group interaction effects did appear in the combined sequence analysis ($F = 15.11$, $df = 4/132$, $P < .01$) and in the analyses of each sequence separately. For Sequence I, in which the groups were all learning List 1, $F = 10.53$, $df = 4/66$, $P < .01$, and for Sequence II, $F = 7.08$, $df = 4/66$, $P < .01$. The appearance of this effect, even when the groups were learning the same list, suggests that it was due to the considerable differences in the ease with which individual paired associates were learned. In Sequence I, the difficult pairs in List 1 were variously classified as Types A, S, and C, according to the interlist relations they would form when the groups subsequently learned the other lists. These differences in learning the individual paired associates were somehow related to the stimulus components of the lists, since certain pairs sharing the same stimulus words were uniformly more difficult to learn in all the lists.

This same interaction also appeared in analyses of the number of trials required to reach the criterion of correctly anticipating all responses of a given type on a single trial ($F = 8.82$, $df = 4/66$, $P < .01$ in the combined sequence analysis). The mean trials required to reach this criterion for Types A, S, and C, respectively, were 8.34, 8.95, and 9.23 in Sequence I, and 8.92, 8.26, and 9.34 in Sequence II. Neither these differences, nor those due to other sources in the analysis, can be considered reliable (all Fs < 1.80). Thus any type differences in performance appearing when these lists were presented in the transfer test can be assumed to be the differential effect of previous learning.

Second list learning.—The possibility that the use of *Ss*' performance in learning the original list as a covariance control might effect greater precision in the analysis of the results of the transfer and recall tests had to be rejected, as the regression coefficients between original list performance and the other measures were uniformly small and could not be regarded as reliably different from zero.

The number of trials required to reach the criterion of correctly anticipating in a single trial all responses in the second lists, regardless of type memberships, was analyzed for overall differences in the learning of the lists. For the groups learning List 1 as their second list in Sequence II, 9.83, 8.67, and 10.50 trials were required, respectively. The groups learning Lists 2, 3, and 4 in Sequence I needed 10.67, 9.00, and 9.00 trials, respectively. These group performances cannot be regarded as different ($F = .58$, $df = 2/66$), nor were the other sources of variance significant.

Differences in verbal generalization related to differences in interlist relations have typically been found to be greatest in the initial stages of second list learning and original list relearning (cf. Haagen, 1943; Morgan & Underwood, 1950; Osgood, 1946, 1948; Underwood, 1951). Therefore, performances on the first anticipation trial on the second lists were examined.

TABLE 3

ANALYSIS OF VARIANCE OF PERFORMANCES IN LEARNING SECOND LISTS

Source	*df*	Correct Anticipations First Trial		Correct Anticipations All Trials		Trials to Criterion	
		MS	*F*	*MS*	*F*	*MS*	*F*
Within *Ss*	144						
Types	2	31.60	43.89**	753.50	48.30**	71.30	10.45**
Types × Seq.	2	1.85	2.57	173.40	10.45**	34.90	5.12**
Types × Groups	4	3.72	5.17**	452.55	29.01**	76.30	11.19**
Types × Groups × Seq.	4	.90	1.25	9.98	.64	1.10	.16
Error 1	132	.72		15.60		6.82	
Between *Ss*	71						
Sequences	1	1.50	1.35	65.50	.16	10.20	.30
Groups	2	2.20	1.98	216.05	.52	50.10	1.46
Seq. × Groups	2	3.05	2.75	117.70	.28	30.30	.88
Error 2	66	1.11		418.22		34.31	

** $P < .01$.

The results of these analyses are presented in Table 3.

A Type × Group interaction effect appeared in this analysis as it had in analyses of performances in the original learning of these same lists. The different types of interlist relations were also an appreciable source of variation.

In order to compare the Type means in the two sequences, Snedecor's (1956) procedure for testing comparisons among all means was followed, using the variance estimate from the combined sequence analysis to establish confidence intervals about the means. This indicated that all differences within and between sequences were significant at the .05 level, except for the difference between the Type A means in the two sequences, and the difference between the Type A and S means in Sequence II, which could not be considered reliable. The Type means for each sequence are presented in Table 4.

In order to examine the effect of the interlist relations beyond the first trials, an analysis of the total correct anticipations produced throughout the course of learning the second lists was performed. In addition, the number of trials required to reach the criterion of correctly anticipating all words of a given type within a single trial was also analyzed. The results of these

TABLE 4

PERFORMANCES ON WORDS OF EACH TYPE IN LEARNING SECOND LISTS

Type	Correct Anticipations First Trial				Correct Anticipations All Trials				Trials to Criterion			
	Sequence I		Sequence II		Sequence I		Sequence II		Sequence I		Sequence II	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
A	1.97	.99	1.78	.97	33.00	11.97	31.42	13.68	5.56	3.82	6.56	4.60
S	1.27	.94	1.56	1.23	30.39	11.08	30.78	12.71	6.64	4.34	6.12	4.09
C	.36	.63	.78	.26	23.72	12.14	28.22	12.23	8.81	3.59	7.23	4.16

TABLE 5

ANALYSIS OF VARIANCE OF PERFORMANCES IN RECALLING AND RELEARNING ORIGINAL LISTS

Source	*df*	Correct Anticipations Recall Trial		Correct Anticipations All Trials		Trials to Criterion	
		MS	*F*	*MS*	*F*	*MS*	*F*
Within *Ss*	144						
Types	2	17.00	20.24**	59.60	4.53*	12.85	3.34*
Types × Seq.	2	.20	.24	4.20	.32	1.20	.31
Types × Groups	4	6.75	8.04**	94.40	7.18**	29.78	7.74**
Types × Groups × Seq.	4	.90	1.07	6.20	.47	2.72	.71
Error 1	132	.84		13.15		3.85	
Between *Ss*	71						
Sequences	1	.00	.00	.00	.00	17.20	1.28
Groups	2	2.60	1.50	156.10	.48	10.90	.81
Seq. × Groups	2	.50	.29	32.05	.10	5.45	.40
Error 2	66	1.73		323.31		13.48	

* $P < .05$.
** $P < .01$.

analyses are also presented in Table 3.

In both analyses the Type effect continued to be significant. Again, the Type × Group effect appeared in each, and the Type × Sequence effect also attained significance in these analyses. The Type means within each sequence in both analyses are presented in Table 4.

Applying Snedecor's procedure to the means of the total correct anticipations produced on all trials revealed that all Type differences within each sequence were significant, except for that between Types A and S in Sequence II. All Type differences between sequences were significant except for those between Type S in Sequence I and Types A and S in Sequence II.

In the analysis of the trials required to reach criterion, all differences within Sequence I were found to be significant, but in Sequence II Types A and S were different from Type C but not from each other. Between sequences, all differences were reliable except for those between Type S in Sequence I and Type A in Sequence II and between the two Type S means.

Recall and relearning of original lists.—The relearning of the original lists was a comparatively easy task for all *Ss*. The numbers of trials needed in Sequence I to reattain criterion on List 1 were 4.75, 5.00, and 6.08 for Groups 1, 2, and 3, respectively, while in Sequence II the groups relearning Lists 2, 3, and 4 required 5.92, 6.17, and 7.50 trials, respectively. The differences between these groups cannot be considered significant ($F = .89$, $df = 2/66$), nor were the other sources of variance significant.

The analyses of the number of correct anticipations for words of each type on the recall trial are presented in Table 5. Both Type and Type×Group interaction effects were found. Within both sequences Types A and S differ significantly from Type C, but not from each other. None of the Type A and S differences between sequences were reliable, but each was significantly different from either Type C means. The type means from both sequences are presented in Table 6.

Performance on relearning the original lists was measured by the total

correct anticipations given on all trials and by the number of trials required to reattain criterion. Analyses of these performances are presented in Table 5. Both Type and Type × Group effects appeared in each analysis. The Type means in each Sequence obtained from analysis of both measures of relearning are presented in Table 6.

In the analysis of the number of correct anticipations produced throughout relearning, it was found that, within both sequences, Types A and S were both reliably different from Type C, but not from each other. As in the recall test, none of the Types A and S differences between sequences were reliable, but each was itself significantly different from either Type C means.

The analysis of the number of trials required to reach the relearning criterion in Sequence I indicated that the Type A and S means were each reliably different from the Type C mean, but not from each other. None of the differences among the Sequence II means were significant. Type A in Sequence I differed from all Sequence II means and Type S in Sequence I differed from Types A and C in Sequence II, but none of the other between-sequence differences were reliable.

Discussion

All analyses of second-list learning and first-list recall and relearning clearly indicate that associative interlist relations are sufficient to produce verbal generalization. Indeed, the results of the most sensitive test of interlist transfer indicate that words participating in associative relations, irrespective of any attendant semantic affinities, were more easily anticipated than words linked only by semantic relations. This suggests that associative relations may even be a necessary factor in interverbal generalization, and that words having semantic relations typically yield generalization effects only because they are usually associated or that they are particularly susceptible to being temporarily associated.

If this were so, the manner in which associative factors operate might be as follows. At the start of learning the pairs in the transfer tests, the responses most likely to occur upon the initial presentations of the stimulus words are those which have been paired to these same words in the initial lists. But because of the establishment of associative connections prior to the experiment, there exists a substantial tendency for these responses, intruding from the preceding lists, to evoke the correct responses in the transfer tests. Hence, these presumably implicit interlist intrusions facilitate the learning of the second list, as their occurrences act as mnemonic aids to learning the transfer list.

TABLE 6

Performances on Words of Each Type in Recalling and Relearning Original Lists

Type	Correct Anticipations Recall Trial				Correct Anticipations All Trials				Trials to Criterion			
	Sequence I		Sequence II		Sequence I		Sequence II		Sequence I		Sequence II	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
A	2.94	1.00	2.81	1.02	16.28	9.27	15.86	9.81	2.86	2.00	3.78	3.06
S	2.69	1.02	2.69	1.10	16.36	9.91	16.25	9.93	3.03	1.74	3.64	3.36
C	1.92	1.16	1.96	1.22	14.36	9.76	14.89	10.95	3.92	2.71	4.08	2.89

This formulation has considerable appeal because it is readily extendable to other generalization situations, and because the mediating processes referred to are exclusively verbal. Thus, they are presumably subject to rather direct assessment and manipulation, and require no supplementary theoretical treatments of semantics or other philosophically redoubtable topics.

In spite of its appeal, this formulation of verbal generalization does not seem to account for the positive transfer found for words participating in the Type S interlist relation in Sequence I, where the required associative connections between lists were apparently lacking. However, this failure may not be in the formulation itself, but in the index of associative connection used in this experiment. The strength of some of the interlist associative connections between the words of Type S in Sequence I may have been considerably greater in the experimental situation than had been inferred from the word association test norms. That is, there may have been features of the experimental situation which produced a temporary increase in the strength of the interlist connections between some of the words of Type S sufficient to account for the limited positive transfer found.

A study reported by Storms (1958) provides evidence for the possibility that temporary supplemental strengthening of associative connections does occur under conditions similar to this. A list of words was read aloud to a group of *S*s who were told they were to be subsequently tested for their retention of these words. The list contained a number of words which appeared with very low frequency as associative responses to a set of stimulus words in an association test (median frequency was 1%). When the association test was administered under these conditions, the median net increase in responding with the words which had recently been heard was 18%. Evidently, these conditions made the usually weak response "more available" for association.

The present experiment provides a situation in which the strength of some of the associative connections between the words of Type S in Sequence I, although very low under most conditions, may have been temporarily supplemented in a manner similar to that found by Storms, for on the first exposure to the transfer list, *S*s were told to read the words aloud and try to anticipate them correctly thereafter. This possibility is empirically meaningful, for it may be possible to discover associative connections that are not susceptible to supplemental strengthening, or to present materials in generalization tests which do not provide the conditions for supplemental strengthening of weak associations.

Summary

Transfer and recall of lists of paired associates were compared in situations in which the original and subsequent lists had identical stimulus words but in which the response words were variously related. The response words in the different lists were either strongly associated but not similar in meaning, highly similar in meaning but not associated, both similar in meaning and associated, or neither similar in meaning nor associated.

The results indicated that positive transfer occurred whenever the words were linked by associative connections, irrespective of semantic relations. Some positive transfer also occurred to a lesser extent across words highly similar in meaning but only weakly associated. It was suggested that the latter result was due to a temporary supplementation of some of the interlist associations connecting these words.

The results suggested a formulation of verbal generalization in which the tendency for the occurrence of the generalized responses is increased because of their associative connections to the responses learned in the original task. On the basis of this formulation, generalization between words judged to be similar in meaning occurs only because associative tendencies connecting them are usually present in *S*s' verbal behavior.

References

Collier, R. O., Jr. Analysis of variance for correlated observations. *Psychometrika*, 1958, **23**, 223–236.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47,** 196–229.

HAAGEN, C. H. Learning and retention as a function of the synonymity of original and interpolated tasks. Unpublished doctoral dissertation, State University of Iowa, 1943.

HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27,** 453–463.

KOGAN, L. S. Analysis of variance: Repeated measurements. *Psychol. Bull.*, 1948, **45,** 131–143.

MCCLELLAND, D. C., & HEATH, R. M. Retroactive inhibition as a function of degree of association between original and interpolated activities. *J. exp. Psychol.*, 1943, **33,** 420–426.

MORGAN, R. L., & UNDERWOOD, B. J. Proactive inhibition as a function of response similarity. *J. exp. Psychol.*, 1950, **40,** 592–603.

OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36,** 299–301.

OSGOOD, C. E. An investigation into the causes of retroactive interference. *J. exp. Psychol.*, 1948, **38,** 132–154.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *ONR tech. Rep.*, 1954, No. 11. (Contract No. N8onr-66216, Univer. Minnesota)

RUSSELL, W. A., & JENKINS, J. J. Associative frequencies for Kent-Rosanoff stimuli and responses under various conditions. *ONR tech. Rep.*, 1956, No. 12. (Contract No. N8onr-66216, Univer. Minnesota)

RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49,** 287–293.

SNEDECOR, G. W. *Statistical methods.* (5th ed.) Ames: Iowa State Coll. Press, 1956.

STORMS, L. H. Apparent backward association: A situational effect. *J. exp. Psychol.*, 1958, **55,** 390–395.

UNDERWOOD, B. J. *Experimental psychology: An introduction.* New York: Appleton-Century-Crofts, 1949.

UNDERWOOD, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first list learning. *J. exp. Psychol.*, 1951, **42,** 44–53.

(Received May 11, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 80–87

THE INFLUENCE OF THIRST AND SCHEDULES OF REINFORCEMENT-NONREINFORCEMENT RATIOS UPON BRIGHTNESS DISCRIMINATION [1]

ROY LACHMAN

Hollins College

Experimental variations of appetitional drive states have produced assorted behavioral effects. The results of research directed toward uncovering the relationship between drive and discrimination processes have been most conflicting (Birch, 1945; Dodson, 1918; Meyer, 1951; Myers, 1952; Tolman & Gleitman, 1949). This may be due in part to certain effects of deprivation time (T_d) that interact with other variables to produce the behavior observed.

Limiting our inquiry to brightness discrimination tasks, it would appear that various drive levels produce specific differences in limited aspects of the response. For example, an increase in T_d will result in, among other things, an increased rate of locomotion. Consider now the effect this will have upon other variables: faster running results in less delay of reinforcement (T_g); also faster running permits the stimuli to impinge on the organism's receptors for a shorter period of time (T_{sd}). Thus, a low-drive producing slow locomotion will result in a longer delay of reinforcement which may in turn slow up the acquisition rate for the choice behavior. In contrast, the longer latencies of low motivation will increase the stimulus duration time (T_{sd}) possibly resulting in quicker learning (Amsel, 1952). The relationship between drive level and $T_{sd} - T_g$ was one of the considerations that generated a "miniature theory," the details of which are presented elsewhere (Lachman, 1960).

The present study seeks to apply the delay of reward, stimulus duration interpretation of drive to the learning of a difficult discrimination. The superiority of high-drive *S*s should be limited to the learning of easy discrimination problems where a shorter delay of reward is produced by high-drive locomotion rates. In a difficult discrimination, the duration of choice-point stimulation is assumed to be critical. Here, the superior delay of reward for high-drive and the superior stimulus duration interval of low-drive might cancel each other. Drive level, therefore, would produce *no* differences in rate of learning. Differences in learning rate between high- and low-drive groups in a difficult discrimination should be reinstated if the choice-point T_{sd} level is equated for the respective drive groups.

Control of choice-point T_{sd} might be accomplished in a Y maze by employing transparent doors and a starting interval of sufficient duration. With transparent doors, gross differences might be eliminated in the duration during which high- and low-

[1] This article is based upon a dissertation submitted to the Department of Psychology, New York University, in partial fulfillment of the requirements for the PhD degree. The author is greatly indebted to his sponsor, H. H. Kendler, for his advice and encouragement throughout the course of this investigation. Thanks are also due M. R. D'Amato for technical assistance. The study was conducted during the author's tenure as a Samuel S. Fels Fellow.

drive groups observe the cue characteristics of the choice point. A facilitative effect of high-drive should occur. If opaque doors are employed in a discrimination problem of the same high difficulty level, differences in learning rate due to drive should not occur.

Comparisons of discrimination attainment between conditions of opaque and transparent doors might be most interesting if forced choice training is employed. The two types of maze door vary the character of choice-point stimulation on forced choice trials and an empirical interpretation might be provided for the forced choice training procedure.

A problem of no minor theoretical importance is the relative influence of trials terminating with the selection of the reinforced and nonreinforced discriminanda alternatives. In the free response situation, the ratio of occurrences of correct and incorrect responses is an uncontrolled consequence of *S*'s behavior.

During discrimination acquisition with a nonpreferred S+ (e.g., the reinforcement of responses to the brighter stimulus for negatively phototropic organisms), the proportion of R+ responses begins in the neighborhood of .25 and terminates at unity. The initial R+ proportion for a preferred S+ is of the order of .75. Differences in acquisition with preferred and nonpreferred discriminanda may involve different learning rates (slopes) as well as unequal initial habit levels. Stated differently, the problem involves the relative influence upon discrimination acquisition of responses to S+ as compared with responses to S−. This problem may be investigated by keeping the initial habit level constant (employing the nonpreferred cue as S+) and varying the response ratio (R+ to S+)/(R− to S−) by forced choice trials.

The investigation thus concerns three areas of influence upon discrimination learning: drive level, choice-point cue availability, and the ratio of responses to the stimulus alternatives.

METHOD

Subjects.—The *Ss* were experimentally naive, male albino rats obtained from the Blue Spruce Farms. At the start of the experiment, they ranged in age from 65 to 85 days. The results are based on data collected from 72 *Ss*, 6 in a cell. Of the 36 low-drive *Ss*, 4 were replaced owing to failure to run.

Apparatus.—The discrimination equipment[2] consisted of completely automatic, single-unit Y mazes, each possessing three identical arms. This arrangement permitted automatic programing of training conditions and dispensed with the handling of *Ss* between successive trials of the same day. The four Y mazes were completely independent and identical in structure, except for opaque vs. transparent doors. Self-contained, electronic programing devices were available for each set of apparatus. The *Ss* were run two or three at a time. The latency and choice data as well as the experimental program for each *S* on every trial were stored in "memory" equipment. Every 5 sec. a scanning device collected the data for each apparatus and transmitted it to a Clary "printout" which recorded the information on a continuous tape.

At the start of the daily training session, *S* was placed in one of the three arms which served, for that trial, as a starting compartment. One of the remaining two arms contained S+, the positive stimulus, while the other one contained S−. After *S* made his choice and completed the trial, the arm which he chose became the starting compartment for the next trial. The other two arms contained S+ and S−. This procedure continued until *S* completed the number of trials programed.

Maintenance (experimental operations for the drive concept).—Two levels of water deprivation were utilized. High-drive *Ss* were watered for 4 hr. on alternate days, 15 min. after being removed from the apparatus. Low-drive *Ss* were watered for 10 min., 4 hr. prior to training, and were administered

[2] The apparatus was designed by M. R. D'Amato (cf. D'Amato & Jagoda, 1960).

a second water ration for 1 hr., 15 min. after training. Since training took place every second day, on nontraining days, low-drive *Ss* were watered for 10 min. followed by an additional 1 hr. of water, administered 5 hr. later. Food was present at all times in the home cage. A temperature of 78° ± 3° and relative humidity varying between 25% and 45% were maintained in the living quarters and experimental rooms.

Discrimination training.—A free choice or forced response to the bright arm (S+) was always rewarded with water; responses to the dark arm (S−) were never rewarded. The starting stimulus was a flashing light less than .2 ft-c in intensity; S+ was a continuous light of 1.6 ft-c.; S− .4 ft-c. The intervals employed were: stimulus, 10 sec. ± .5 sec.; intertrial, 25 sec. ± 2.5 sec. (lights out); starting, 8 sec. ± .5 sec. The water reinforcement duration was 2 sec. ± .5 sec., i.e. *S* was permitted to drink for an average of 2 sec. after making physical contact with the water.

Forced trials were employed to produce ratios of correct responses to incorrect responses of 3/1, 1/1, and 1/3. This was accomplished by programing blocks of 4 trials, 1 free and 3 forced. One of the forced trials was programed as contingent upon the free trial; the contingent trial always produced a response which was opposite to the free trial response. The remaining two forced trials were either both positive, or both negative, or one positive and one negative; the ratios stated were thus produced. Since responses to the brighter stimulus received 100% reinforcement and those to the darker stimulus received 0% reinforcement, *Ss* with a correct/incorrect ratio of 3S+/1S− ended each interval of 4 trials with 75% of their training trials to the reinforced arm (Group 75%). Likewise, for each block of 4 trials, the 1S+/1S− and 1S+/3S− ratio groups were administered forced choices that resulted in 50% (Group 50%) and 25% (Group 25%) of their trials, respectively, being run to the correct stimulus. On the final day of training, all 40 trials were free. The final test trials were employed because postremity effects were suspected to be operating in the forced trial phase of training. That is, the response on a free trial may have been influenced by the direction of forcing on the immediately preceding forced trial. Furthermore, the free trials of the final day eliminated the long latencies typically encountered during the terminal stages of forced choice training (Spence, Goodrich, & Ross, 1959).

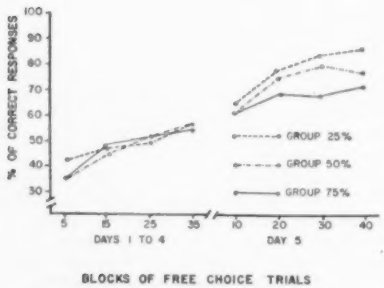

FIG. 1. The percentage of correct responses on free choice trials for forced choice response schedule groups.

TABLE 1

ANALYSIS OF VARIANCE FOR THE NUMBER OF CORRECT RESPONSES ON DAY 5

Source	*df*	*MS*	*F*
Schedules (S)	2	108.10	3.50*
Drive (Dr)	1	36.13	1.17
Doors (Do)	1	342.35	11.08**
S × Dr	2	35.78	1.16
S × Do	2	1.26	<1
Do × Dr	1	1.12	<1
S × Dr × Do	2	34.13	1.10
Within	60	30.89	
Total	71		

* $P < .05$.
** $P < .01$.

All *Ss* were given 20 trials on Day 1 and 40 trials a day on Days 2 to 5. A detailed account of forcing procedure and preliminary training will be found in Lachman (1960, Exp. 2).

Experimental design.—A 2 × 2 × 3 design was adopted: High (H) and low (L) thirst drive; transparent (T) and opaque (O) maze doors; and three ratios of response, 3/1 (75% choice of S+), 1/1 (50% choice of S+), and 1/3 (25% choice of S+).

RESULTS

Figure 1 shows the learning curves for the three response schedule groups. The curves begin at below-chance level because of *Ss*' preference for S− (dark). By Day 4, differences are still absent and all groups are responding slightly above the chance

level. The exclusively free choice trials of Day 5 exhibit divergent curves. During the terminal 10 trials of Day 5, Group 25% is responding close to the 90% correct response level, Group 75% is performing near the 70% level, and Group 50% is between the two.

Table 1 reports the analysis of variance for the number of correct responses on Day 5; Table 2 reports the individual cell means. The analysis produced two significant effects: schedules of forced choice responses and transparent-opaque doors. The average number of correct responses on Day 5 is 31.42 for Group 25% (i.e., the group trained with 25%

TABLE 2

CELL MEANS: THE AVERAGE NUMBER OF CORRECT RESPONSES ON DAY 5

Forced Response Schedules	High Drive		Low Drive	
	Transparent	Opaque	Transparent	Opaque
25%	33.5	28.8	34.0	29.3
50%	31.8	30.0	29.7	23.3
75%	31.2	23.8	27.8	26.0

of R+ and 75% of R−), 28.83 for Group 50%, and 27.21 for Group 75%. For transparent and opaque doors the correct response averages are, respectively, 31.33 and 26.97.

The differences in learning rate for drive levels are not significant. The high-drive *S*s made an average of 29.86 correct responses on Day 5, the low-drive *S*s 28.44. No significant interactions were obtained.

Figure 2 graphically reveals the reason for the significant door effect. Group T (transparent doors) started and remained at a higher level of response throughout the training days of intermixed free and forced choice trials. On the totally free choice trials of Day 5, the performance of

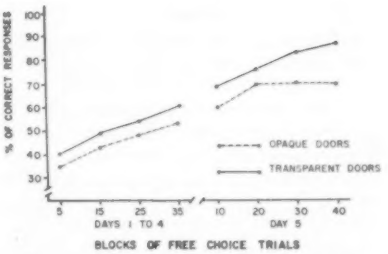

FIG. 2 The percentage of correct responses on free choice trials for transparent and opaque conditions.

Group T continues to improve with terminal responses at the 87% correct-response level. By the second block of 10 trials for the same day, Group O levels off at the 70% level and remains there until the termination of training.

Thus the choice data show that transparent doors in a combined free and forced choice situation facilitate discrimination learning and that the speed of this learning tends to increase with greater proportions of negative instances (incorrect, unrewarded, forced choices).

The results for the running data are presented in Fig. 3. The percentage of trials completed within

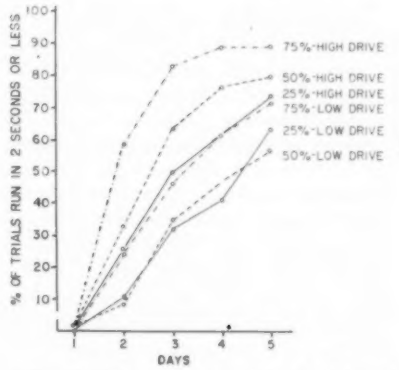

FIG. 3. The percentage of free and forced trials (Days 1 to 4) and free trials (Day 5) run in 2 sec. or less for drive and response schedule groups.

TABLE 3

CELL MEANS: THE AVERAGE NUMBER OF TRIALS COMPLETED WITHIN 2 SEC. FOR ALL TRAINING DAYS

Forced Response Schedules	High Drive		Low Drive	
	Transparent	Opaque	Transparent	Opaque
25%	84.17	85.83	67.67	51.67
50%	110.83	94.67	62.33	57.33
75%	136.00	122.17	80.17	86.33

2 sec. is plotted as a function of training days. Forced choice response groups for each of the drive levels are plotted as parameters. Each curve represents the running data for 12 *S*s since transparent-opaque cells were collapsed for graphic presentation. Figure 3 indicates that rate of running increases with the proportion of forced choices to the rewarded stimulus and with drive level. Group H-75% improves its locomotion rate in a highly accelerated fashion. The running time of Group H-50% is likewise facilitated but to a lesser degree. Group L-75% is facilitated in running speed to a degree that eliminates running differences between it and Group H-25%.

Since running times of less than 1 sec. were recorded as 1 sec., the large number of 1-sec. scores precluded the utilization of logarithmic or reciprocal transformations. As a result, the number of trials completed in 2 sec. or less was computed for each *S*. Enumeration data, thus obtained, more closely approximated analysis of variance assumptions. "Percentage of trials completed within 2 sec." also provided the best single descriptive statistic. Correlations computed between the 2-sec. trial measure and more conventional measures of running time were higher than the intercorrelations of the conventional locomotion scores (Lachman, 1960, Exp. 1.)

The running time averages for each cell is presented in Table 3. The analysis of variance based on these data is presented in Table 4. The score for each *S* is based on the total number of its trials completed with a running rate of 2 sec. or less. Forced choice schedules and drive level are both highly significant. No interactions are significant.

DISCUSSION

The results here obtained severely strain the credibility of the stimulus duration, delay of reward, drive model. Most damaging for the model are the locomotion data. Although both high drive and greater proportions of forced choice trials to the rewarded stimulus markedly accelerate locomotive behavior, only the latter influences choice behavior. If differences in delay of reward and choice-point stimulus duration are functionally related to discrimination acquisition, then acquisition effects should occur regardless of the ultimate source of the differential reward delay and cue duration. Response ratio schedules and thirst drive have similar effects on running rate and presumably upon the delay and duration variables; similarities in the acquisition of choice behavior should result. Clearly, this is not the case. Forced choice response schedules influenced both running time and discrimination learning. An increasing propor-

TABLE 4

ANALYSIS OF VARIANCE FOR THE NUMBER OF TRIALS COMPLETED WITHIN 2 SEC.

Source	*df*	*MS*	*F*
Schedules (S)	2	7,374.85	11.40***
Drive (Dr)	1	26,030.01	40.24***
Doors (Do)	1	931.68	1.44
S × Dr	2	737.93	1.14
S × Do	2	68.34	<1
Dr × Do	1	91.13	<1
S × Dr × Do	2	582.04	<1
Within	60	646.87	
Total	71		

*** $P < .001$.

tion of positive instances (rewarded forced choices of S+) increased the acquisition of the running response and perhaps its asymptote while simultaneously retarding the acquisition of the discrimination habit. In contrast, thirst level produced no influence upon the measure of discrimination learning, although it significantly influenced locomotion rate. These findings support the conclusion that discrimination learning is, to a degree, independent of the concurrent locomotive behavior and a concurrent or multiple habit hypothesis may be warranted.

Asymmetry of forced responses to the rewarded and negative stimuli poses an interesting problem. Group 75% ran three-fourths of its training trials to S+. Thus, this group received both a greater number of trials *and* a greater number of reinforced trials to the positive stimulus as compared with Group 25%. If a simple reinforcement model assumes no influence of negative (nonrewarded) instances, then according to either a reinforcement or a contiguity hypothesis Group 75% should have exhibited superior discrimination learning. The opposite was observed. It follows that *neither* a simple reinforcement nor contiguity theory can account for this result. The attempt to reduce the acquisition rules for discrimination learning to a single learning principle would appear to involve an oversimplification. Special rules of acquisition appear to be necessary in order to account for the choice data of discrimination learning. In contrast, a simple reinforcement assumption adequately describes the acquisition of the running response.

The present observations are consistent with the results of Denny and Dunham (1951) with rats and Cantor and Spiker (1954) with preschool children. These two investigations, unlike the present study, kept the number of positive instances constant and varied the number of negative instances.

The possible aversive effect of negative instances is by no means a novel formulation and has been considered, among others, by Spence (1958, p. 101). He postulated fractional antedating frustration reactions (r_f) for nonrewarded choice responses. Perhaps an r_g—r_f model is required to account for the present data. According to an r_g—r_f theory, the response schedule data (superiority of negative instances) would be interpreted as demonstrating a steeper acquisition gradient for r_f as compared with the acquisition of r_g. That is, classical conditioning of an avoidance response to the nonrewarded stimulus would appear to occur at a faster rate than the conditioning of r_g to the positive stimulus. If the correct choice response were assumed to increase in strength with higher levels of both r_g and r_f, the superiority of Group 25% would reflect the greater degree of conditioning of r_f.

It will be recalled that forced choice training was terminated *prior* to the appearance of difference in locomotion rate to S+ and S−. On Day 5, all trials were free choice. The locomotion rates on Day 5 for the several experimental groups maintain the same relative order as on the previous days when forced choice training was administered. Consequently, the graphic presentation and analysis of locomotion rates was based on data for all trials: free, correct forced choice, and incorrect forced choice.

Continuing the analysis according to Spence's theory, the superior locomotion rate of the 75% groups might be interpreted as reflecting a higher quantitative level of the intervening variable K. The slower running rate for the 25% groups could be interpreted as reflecting a greater level of I or a smaller relative level of K. In either case, the Hull-Spence assumption that "locomotion and choice behavior are equivalent indices of reaction potential" is in trouble. The results for the 75% and 25% groups indicate an inverse relationship between level of discrimination learning and locomotion rate under the initial conditions of the present study. Group 75% exhibited a greater rate of locomotion and an inferior level of discrimination learning. The reverse was observed for Group 25%. Actually, separate mathematical

expressions are required to represent accurately the acquisition of choice and of running.

If we turn for a moment to the thirst data, a more serious difficulty arises for the hypothesis of locomotion and choice behavior as equivalent indices of reaction potential. In addition, the thirst data require a modification of the Hull-Spence assumptions concerning the interrelations between the intervening variables. Drive level produced a highly significant effect upon locomotion while at the same time producing no observable influence upon choice behavior. It follows that although the argument, $E+ = D \times H+$, is a satisfactory hypothesis for the running data, the expression, $(E+)\text{-}(E-) = (H+)\text{-}(H-)$, is more consistent with the choice behavior of the present research. There is an alternative manner by which Hull-Spence theory may accommodate the present effects of thirst. The assumptions concerning the coordinating definitions of D could be modified. That is, hunger and thirst manipulations may be assumed to have *different* influences upon D. This possibility warrants serious consideration.

The results produced by differential accessibility of choice-point cues resulting from the opaque and transparent doors require some explanation. Discrimination learning was significantly facilitated by the transparent doors. The transparent condition permitted additional reception of choice-point cues during the starting interval and observation of stimuli from the unavailable arm during the running of a forced trial. Thus, the enhancement effect may have occurred at either locus. A preliminary study (Lachman, 1960, Exp. 1) involved transparent doors without forced trials. Since the transparent doors did not enhance the discrimination in this preliminary study but did so in the present study, the locus of this effect appears to be at the choice point during forced trials. The argument, $H = f(N)$, might be refined so that N is defined not solely as the number of training trials but includes a term for the proportion of stimulus information available during training.

The discussion, thus far, has concerned theoretical approaches that include the noncorrection training procedure within their boundary conditions. Noncorrection maze learning has generated considerable difficulty for Estes' theory (1959, p. 416) especially in relation to the effects of nonreinforcement. The present results suggest that at least early in acquisition, nonreinforced responses have important consequences. A greater relative increment in the probability of R+ occurs after a response to S− than after a response to S+.

With respect to the model that initiated this research, it must be concluded that the model cannot be evaluated as true or false (or more appropriately as promising or unpromising). It is possible that the model is valid but the technique (thirst) used to evaluate it is not appropriate. The present data potently imply that the model may be invalid. To clearly demonstrate either of these two possibilities, more research is needed, perhaps with hunger instead of thirst.

Summary

A model was proposed to explain the role of drive in discrimination learning. An experiment designed to test the model was conducted with three schedules of forced choice responses. That is, discrimination training was conducted with three different ratios of forced responses to the rewarded and nonrewarded stimuli.

The results and conclusions follow: (*a*) The research failed to produce the outcomes predicted by the model. (*b*) Discrimination learning was decreased by a relatively large proportion of forced choices to the *rewarded* cue and enhanced by a greater proportion of forced choices to the *nonrewarded* cue. (*c*) Level of thirst failed to influence the choice behavior in discrimination learning. However, thirst level exhibited a sizable influence upon running behavior. (*d*) Discrimination learning was improved by the presence of *both* relevant cues at the choice point during forced trials.

Several alternative interpretations of the results were explored, including the implica-

tions for the learning theories of Spence and of Estes.

REFERENCES

AMSEL, A. Rate of learning of a visual brightness discrimination as a function of discriminanda duration. *J. comp. physiol. Psychol.*, 1952, **45**, 341–346.

BIRCH, H. G. The role of motivational factors in insightful problem-solving. *J. comp. Psychol.*, 1945, **38**, 295–317.

CANTOR, G. N., & SPIKER, C. C. Effects of nonreinforced trials on discrimination learning in preschool children. *J. exp. Psychol.*, 1954, **47**, 256–258.

D'AMATO, M. R., & JAGODA, H. Effects of extinction trials upon discrimination reversal. *J. exp. Psychol.*, 1960, **59**, 254–260.

DENNY, M. R., & DUNHAM, M. D. The effect of differential nonreinforcemrnt of the incorrect response on the learning of the correct response in the simple T maze. *J. exp. Psychol.*, 1951, **41**, 382–389.

DODSON, J. D. Relative values of reward and punishment in habit formation. *Psychobiology*, 1918, **1**, 231–276.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

LACHMAN, R. Drive level and forced choice response schedules in brightness discrimination. Unpublished doctoral dissertation, New York University, 1960.

MEYER, D. R. Food deprivation and discrimination reversal learning by monkeys. *J. exp. Psychol.*, 1951, **41**, 10–16.

MYERS, T. I. An experimental investigation of the effects of hunger drive upon the brightness discrimination learning of the rat. Unpublished doctoral dissertation, State University of Iowa, 1952.

SPENCE, K. W. Behavior theory and selective learning. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1958.* Lincoln: Univer. Nebraska Press, 1958.

SPENCE, K. W., GOODRICH, K., & ROSS, L. E. Performance in differential conditioning and discrimination learning as a function of hunger and relative response frequency. *J. exp. Psychol.*, 1959, **58**, 8–16.

TOLMAN, E. C., & GLEITMAN, H. Studies in spatial learning: VII. Place and response learning under different degrees of motivation. *J. exp. Psychol.*, 1949, **39**, 653–659.

(Received May 13, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 88–94

GENERALIZATION GRADIENTS IN A DISCRIMINATION SITUATION [1]

DAVID LABERGE

University of Minnesota

The purpose of this study was to investigate the shapes of human generalization gradients in an effort to obtain information basic to a formulation of a set-model for stimulus generalization.

Gradients were obtained by first training *S*s to respond selectively to two different stimuli on a metathetic dimension (Stevens, 1957), and then testing them on new stimuli in the neighborhood of the training stimuli. The resulting gradients, plotted either in terms of relative frequency or latency, were expected to resemble one side of the traditional generalization gradient (Hull, 1939) which has a peak at the training stimulus and decreasing slopes on both sides of the peak.

The stimuli employed in the present experiments were positions of a small light along a vertical dimension. Since meaningful interpretations of shape characteristics of gradients required knowledge of the psychological spacing of the stimuli, preliminary scaling experiments were carried out to determine the optimal range of stimuli within the dimension to be used for generalization testing.

METHOD

Subjects.—The *S*s were men and women undergraduates from the University of Minnesota. The preliminary scaling experiment employed 16 *S*s and the generalization experiment employed 64 *S*s in the final analysis. All *S*s were naive with respect to experiments of this kind.

Apparatus.—The visual display was located 14 ft. in front of the *S*'s chair. The stimulus, a rectangle of light $\frac{1}{16}$ in. by $\frac{7}{16}$ in., could be positioned precisely along a vertical axis. In order to stabilize the perceived position of the light in the darkened room, two white enameled guide strips, $1\frac{1}{4}$ in. by 68 in., ran the length of the vertical axis on either side and adjacent to the rectangle of light. The lower border was 6 in. above the floor and the upper border was 76 in. above the floor. The position of a stimulus along the dimension was designated according to its distance in inches from the lower border. Thus Stimulus 35 was the physical midpoint of the range of stimuli bordered by the ends of the two white guide strips. Generous use of black commando cloth and flat black paint virtually eliminated all other visual reference points.

The *E* sat behind a light-proof screen and adjusted the position of the stimulus with the aid of a finely graduated scale. The rectangle of light was not illuminated during the time when *E* changed its position, and the room was sufficiently dark so that *S*s could see the rectangle only when it was illuminated.

A response box was located on a table in front of *S* and contained two levers, $5\frac{3}{4}$ in. apart, with red reinforcement lamps located $2\frac{9}{16}$ in. above each lever. An ATC timer[2] controlled the exposure times of the stimuli and reinforcements as well as all interevent intervals. A Beckman counter[3] measured response latencies in .001-sec. units, and a Beckman digital recorder printed out automatically the response latency, response alternative, and reinforcement alternative on each trial.

Scaling experiment.—The scaling procedure was based on measurement of difference thresholds. The simplest technique seemed to be to find a range of stimuli over which equal physical intervals would correspond to equal psychological intervals. The method of constant stimuli was used to obtain the

[1] This research was supported by a grant to the author from the National Science Foundation. Equipment was purchased through a grant from the Graduate School of the University of Minnesota.

[2] Automatic Timing & Controls, Inc., King of Prussia, Pennsylvania.

[3] Beckman Instruments, Inc., Richmond, California.

threshold curves around two standard stimuli, 41 and 45, which were near the border of a range of stimuli which previous studies had indicated as having equal and stable jnd's.

The procedure was to present a series of stimuli, the first of each pair being the standard, and the second a variable stimulus +2, +1, 0, −1, or −2 in. from the standard. The *Ss* were instructed to determine if the second stimulus of each pair was above, below, or the same as the first stimulus. The pairs were randomized in 30-trial blocks, each block containing 6 of each of 5 possible standard-variable pairs. Each *O* was given 8 blocks, totaling 260 trials. The two standards were assigned to successive pairs of blocks in the orders 41, 45, 45, 41 or 45, 41, 41, 45. There were 1-min. rest periods between each pair of blocks. Five pairs of stimuli were given before each block to orient *S* to each new standard. There were four different random sequences, four *Ss* to a sequence. The 16 *Ss* were run two at a time, and were instructed to write their judgments on a tablet of paper. They were also instructed to look down at their hands immediately following the second stimulus of each pair in order to attenuate an autokinetic-type of effect found when *Ss* focused rigidly on the spot where the last stimulus appeared.

The time relations of the constant stimulus experiment were as follows: A buzzer sounded for 1 sec.; time between the buzzer and the standard stimulus, 1 sec.; duration of the standard stimulus, 2 sec.; time between the standard stimulus and variable stimulus, 4 sec.; duration of the variable stimulus, 2 sec.; time between the variable stimulus and the buzzer of the next trial, 4 sec.

Since the psychophysical curves obtained for each of the two standards were virtually identical, we may conclude that within the range of Stimuli 41–45, and probably even within the greater range of Stimuli 39–47, equal physical intervals correspond to equal psychological intervals. Therefore we may reasonably expect to determine shape characteristics of generalization gradients obtained over this range of stimuli.

Procedure.—For all generalization conditions *Ss* were given sufficient training trials with two stimuli, S_1 and S_2, to bring responding at or near asymptote. Then a series of generalization test trials to neighboring stimuli followed, with training stimuli interspersed among them to maintain discrimination performance at asymptote. Reinforcing lights always followed training stimuli such that one response, A_1, was reinforced 100% of the time when S_1 occurred, and the other response, A_2, was reinforced 100% of the time when S_2 occurred. Neither reinforcing light followed occurrences of test stimuli.

There were three discrimination training intervals: Stimuli 42–44, 41–45, and 27–59, which were labeled the 2-, 4-, and 32-in. training intervals. All training intervals had a common midpoint, namely Stimulus 43. Sixteen *Ss* were assigned to each of the three discrimination conditions and were given generalization tests on seven stimuli lying between the training stimuli, S_1 and S_2. In addition, eight *Ss* were assigned to the 2-in. and 4-in. training intervals and were given generalization tests to seven stimuli, six lying outside the training interval and one lying midway between the training stimuli.

The stimuli were randomized without restrictions within blocks of 17 trials. Each *training* block contained eight S_1, eight S_2, and one test stimulus, 43. Each *test* block contained five S_1, five S_2, and seven different test stimuli.

One *S* was tested at a time, and given 35 blocks of trials, totaling 595 trials. The blocks were given in five series, seven blocks to a series, with 3-min. rest periods between each series. The first series contained seven training blocks. Subsequent series contained an initial training block followed by six test blocks. A different sequence of stimuli was constructed for each subgroup of four *Ss*. Side of the reinforcing light was counterbalanced against S_1 and S_2 within each sequence.

Time relations within a trial were as follows: Duration of the stimulus light, 2 sec.; time between stimulus and reinforcing light, 1 sec.; duration of the reinforcing light, 1 sec.; time between reinforcing light and next trial, 3 sec.

The *Ss* were instructed to press a lever when a white light appeared in the front of the room. They were told that a red light above one of the levers would flash to indicate which lever they should have pressed on that trial. In addition, they were told that sometimes no reinforcing light would follow a white light and that such trials were to be considered as test trials wherein *E* wanted to know which lever *S* thought should be pressed on that trial.

Results and Discussion

Discrimination training.—Discrimination curves for *Ss* who learned the discrimination and were given generalization tests are shown in Fig. 1.

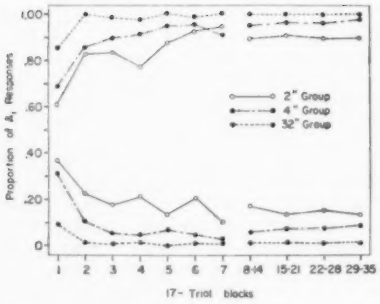

FIG. 1. Discrimination curves for *Ss* over all trials, 16 *Ss* in each group. (The upper and lower curves represent responses to training stimuli S_1 and S_2, respectively.)

The learning criterion was 13 or more correct responses in Block 7 of training. The data from the 2-in. and 4-in. groups which were given generalization tests to the outside are not included here, although their data were the same as the 2-in. and 4-in. groups which had tests to the inside of the training intervals. Data from training stimuli of training and test blocks were combined in Fig. 1, since none of the t tests of differences between test and training blocks approached significance. There are clear separations of the three discrimination groups, with an F test of the differences among asymptotes being significant beyond the .001 level. The curves of discrimination learning for all *Ss*, learners and nonlearners, show even greater separations, and all groups appear to have reached their asymptotes by Blocks 5 or 6. There is no indication that the obtained orderings of group asymptotes will eventually reverse, although a reversal is predicted by a Markov model of discrimination learning proposed by Atkinson (1958). Test trials apparently do not reduce the level of discrimination performance as indicated by testing the differences between the last two blocks of the first series and the seven blocks of the final test series for all groups, including the groups having generalization tests to the outside of the training interval.

Generalization tests.—Since we were interested in generalization gradients

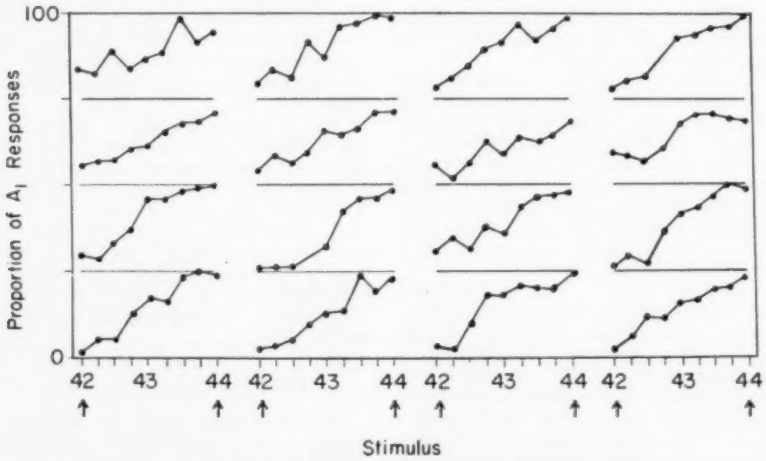

FIG. 2. Individual gradients for the 2-in. training group, where testing was made inside the training interval.

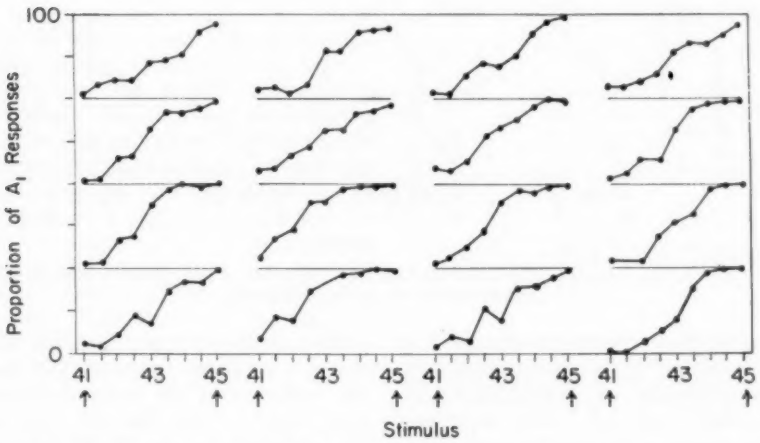

FIG. 3. Individual gradients for the 4-in. training group, where testing was made inside the training interval.

only of *S*s who learned the discrimination, we did not use generalization data from the seven *S*s of the 2-in. condition and five *S*s from the 4-in. condition who did not learn the discrimination in the first seven training blocks.

Individual gradients for each of the discrimination conditions are shown in Fig. 2, 3, 4, 5, and 6. For these curves, each training-stimulus point is based on 120 observations, and each test-stimulus point is based on 24 observations. The curves were obtained by averaging across all test blocks, a procedure justified by the observation that series by series comparisons of group generalization curves

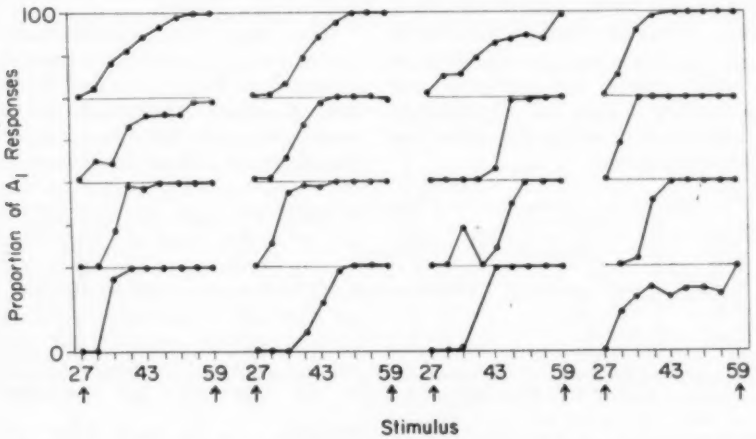

FIG. 4. Individual gradients for the 32-in. training group, where testing was made inside the training interval.

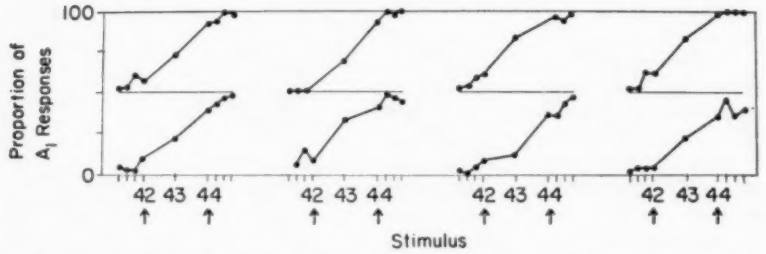

FIG. 5. Individual gradients for the 2-in. training group, where testing was made chiefly outside the training interval.

showed no systematic change. It appears that fairly orderly generalization gradients are obtainable on individuals under all conditions tested. Meaningful interpretations of shape characteristics of the gradients are justified only in the case of the 2-in. and 4-in. inside groups, where the test stimuli fall within the range of values scaled by the preliminary psychophysical experiment.

A notable feature of the individual gradients is their almost linear slopes, which might be expected in cases where stimuli are equally spaced. There is no consistent indication of plateaus near the middle of the gradients, even in the case of the 32-in. group. However, a strong case can be made for the existence of plateaus at either end of the gradients near the training stimuli for certain individuals who learned the discrimination to a high criterion.

The asymmetry of the 32-in. gradients is probably due to the fact that the midpoint of the vertical dimension, namely Stimulus 35, was not the midpoint of the 27–59 training interval. And verbalizations of *Ss* confirmed that the dimension's physical midpoint served as a strong reference point for most judgments of the 32-in. condition.

In the cases of the generalization tests to the outside of the S_1-S_2 interval shown in Fig. 5 and 6, there is a shift in the peak response rate away from the training stimuli, an effect similar to that found in the pigeon studies of Guttman and Kalish (1956) and Hanson (1959).

The slope characteristics of the group gradients may be compared by referring to Fig. 7, 8, and 9. Statistical tests of differences in slopes were prepared by first computing the slopes of a least-squares straight

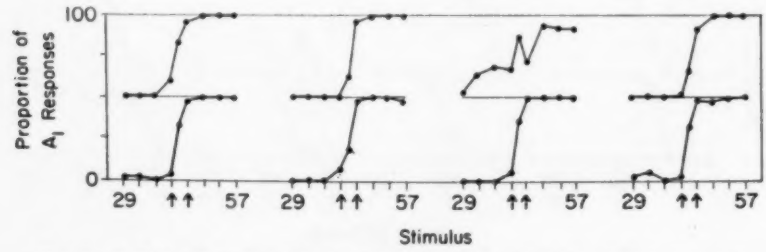

FIG. 6. Individual gradients for the 4-in. training group, where testing was made chiefly outside the training interval.

line fit to each *S*'s data. Because of plateau effects near training stimuli, some of the points in these regions were not used in the least-squares analysis. The criterion for selecting points was as follows. Reading the graphs from left to right, we took the last stimulus point whose response proportion was less than .10 and the first stimulus point whose response proportion was greater than .90, plus all stimulus points between these two points. The resulting distributions of individual slopes had little overlap and *F* tests of differences among

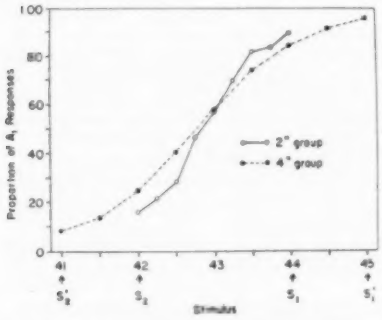

FIG. 7. Group frequency gradients for the 2-in. and 4-in. training groups, where testing was made inside the training intervals. (Each curve represents data of 16 *Ss*.)

slopes for the three discrimination intervals were highly significant. For both the 2-in. and 4-in. conditions, the mean slopes of the gradients obtained from testing stimuli inside the training interval were significantly greater than the mean slopes of gradients obtained from testing stimuli outside the training interval. There was also a significant difference between the response lever associated with the upper and lower training stimulus.

The exact form of the individual generalization gradient could not be determined from the present data, in

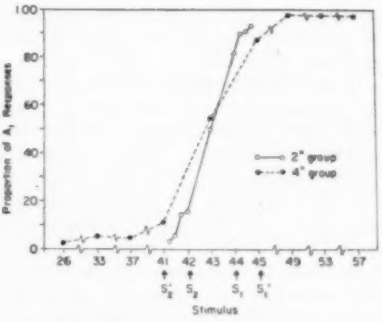

FIG. 8. Group frequency gradients for the 2-in. and 4-in. training groups, where testing was made chiefly outside the training interval. (Each curve represents data of 8 *Ss*.)

spite of efforts to provide careful measurements of stimulus spacings for the 2-in. and 4-in. groups. The difficulty stemmed from the fact that all curves, including individual curves, were averages of more basic gradients. The sigmoid shape of the group curves of Fig. 7 can be obtained from averaging linear curves which have plateaus of varying length near the training stimuli. And similarly, an individual gradient would assume a sigmoid shape if fluctuations in acuity during the testing series produced plateaus of varying length. Until a method is found for obtaining generalization gradients under rigidly

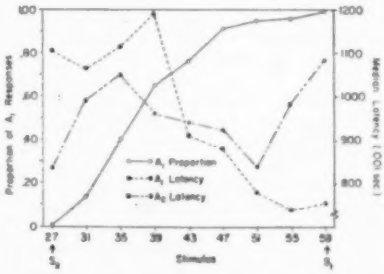

FIG. 9. Group frequency and latency gradients for the 32-in. training group, where testing was made inside the training interval. (Response A_1 was reinforced when S_1 appeared and response A_2 was reinforced when S_2 appeared. Each curve represents data of the same 16 *Ss*.)

constant conditions, the determination of the exact shape of the generalization gradient will remain an elusive prize.

Group latency gradients plotted within the training interval indicate a monotonic increase in latency as the test stimulus is removed farther from the training stimulus, with the curves of Fig. 9 representing the poorest orderings obtained. All other group gradients are much more orderly, showing the same crossing of the two response curves such that the incorrect response is always slower than the correct response. For the cases where tests were made outside the training interval, the group curves continued their monotonic form over the range of outside test stimuli, with the fastest latencies being associated with the test stimuli outside the training stimuli.

Summary

Sixty-four *Ss* were given 119 trials on a discrimination problem involving two different positions of a light on a vertical dimension. Following this they were given 476 trials made up of (*a*) generalization tests to neighboring stimuli along the dimension; and (*b*) further training trials with reinforcements to maintain asymptotic discrimination performance during testing.

Individual and group frequency generalization gradients were fairly orderly. The magnitude of the interval between training stimuli had a highly significant effect upon the slope characteristic of the gradients. Peaks of the gradients were shifted toward the outside of the interval between training stimuli whenever response rates to the training stimuli were below maximum. Group latency gradients indicated similar monotonic orderings under all conditions.

Discrimination asymptotes were ordered in the expected way with no appreciable change in asymptotic level over the generalization test trials.

References

Atkinson, R. C. A Markov model for discrimination learning. *Psychometrika*, 1958, **23,** 309–322.

Guttman, N., & Kalish, H. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51,** 79–88.

Hanson, H. M. Effects of discrimination training on stimulus generalization. *J. exp. Psychol.*, 1959, **58,** 321–334.

Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46,** 9–30.

Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64,** 153–181.

(Received May 18, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 95-96

SUPPLEMENTARY REPORT: EFFECT OF PRACTICE ON AN ILLUSION [1]

DOROTHY RETHLINGSHAFER AND THOMAS I. SHERRER

University of Florida *Florida State Employment Service*

In a recent study, Rechtschaffen and Mednick (1955) found that *Ss* in darkness would experience autokinetic movement that reflected the suggestion that a pinpoint of light would write words. Wide individual differences were noted, but no practice effects were studied. The present investigation is an attempt to determine the practice effect, if any, in the autokinetic illusion when *Ss* are given the instructions that the light will write English letters and words.

Method.—The *Ss* in the major experimental group were 40 male college students (age range = 17 to 28 yr). Nine men and two women from the same population were used in a control group.

Each *S* was tested on three 20-min. sessions on different days. During the first 10 min. of each session *Ss* were told that the small pinpoint of light would write letters. A rest period, approximately 5 min., was given before a final 10 min. when *Ss* were told that the light had been rearranged so that it would write words.

After first determining whether the illusion of movement was perceived, the first instruction was given:

> The first 10 min. part of the test will be essentially this: the light has been set up so it will trace the outlines of the letters of the alphabet. The letters will vary in size, and will be given in continuous series, with no breaks between them. As soon as you can make out the letters, or distinguish what they are, report them to me as you go along. There will be both capital and lower case letters given in manuscript form just as though I were writing them on a blackboard for you to read. Okay?

After the rest period, the following instructions were given:

> The second part of the test is essentially the same as the first, with the exception that the letters have been arranged so that words will be spelled. We aren't accustomed to using individual letters in reading, writing, or in speaking; for this reason it is believed that it will be easier for you to group the letters into familiar words, rather than to break the series into individual letters. There will be no breaks between words, so this is primarily a test to see if you can tell when one word is complete and another begins. If you make out some of the letters and miss some, you may be able to guess the whole word by the order in which the letters are given. As soon as you have completed the word tell me what it is. Okay?

The *Ss* were seated in a dark room, approximately 12 ft. from the pinpoint of light. The *E* sat near the box from which the light was emitted and recorded responses. A luminous dial watch was used to record time.

The data reported were obtained from two groups. In Group 1, *Ss* made their reports in Sessions 1 and 2 alone and in Session 3 with another *S*. In Group 2, *Ss* were with another *S* in Sessions 1 and 2, and alone in Session 3.

Ten control *Ss* were tested alone with the same instructions as used in Session 1, *but no pinpoint of light was present.* No control *S* reported perceiving letters or words. Five reported slight pinpoints or "flashes" of light which disappeared when attempts were made to focus upon them. The results are reported only for the experimental *Ss*.

Results.—The experimental *Ss* increased in the frequency of their reports throughout the three sessions. Practice affected the perceiving of the autokinetic illusion in the particular form induced by the suggestion given. As seen in Table 1 there was a greater frequency of responses in Session 2 as compared with Session 1. This was true for the reports of both letters and words, and for both groups. Except for reports of letters by Group 1, there was a further rise in frequency of responses in Session 3. Tests of significance of the mean differences indicated that the increases from Sessions 1 to 3 were statistically significant ($P < .05$) except for Group 1 in their reports of letters. For Group 1, the rise in frequency of reports of letters from Session 1 to Session 2 was significant ($P < .05$). The between-sessions F was significant at beyond the .01% level for both groups on words and letters except for the Group 1 words measure, which was significant at beyond the .05 level.

The number of responses of letters was 1005 with all letters of the alphabet being reported. The letter most frequently perceived was *e*, with *m* and *o* being next. There were 238 words perceived, most of them being fairly simple and short words as: *the*, *be*, *of*, *all*, etc. Almost half of the words were reported only once, with a few of these being individual, such as, *Robert*, *nonsense*, *bombs*, etc. Some *Ss* accepted the directions readily. They also shifted easily from the set to perceive letters to perceiving words. At the other extremes were *Ss* who questioned the instructions. They gave few responses and with qualifications.

Certain *Ss* repeated the same responses; others experienced unusual "writing," such as perceiving the letter as being written on its side, or as movement from left to the right, making the writing difficult to perceive.

It is important to note that *Ss* must accept the suggestion that the light is writing. When graduate students were used as *Ss* in a preliminary study, the number of responses were few or nonexistent. The *Ss* used in this report were naive in respect to the illusion but some were more critical than others of the suggestion that the light was writing.

In this study, the particular form of the illusion was induced in *Ss* from the instructions given by *E*. As in the Atkinson and

[1] This study was reported, in part, at the September 1957 meeting of the American Psychological Association.

TABLE 1

EFFECT OF PRACTICE ON AN ILLUSION

Session	Letters				Words			
	Group 1		Group 2		Group 1		Group 2	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	8.10	5.46	5.35	3.29	1.55	2.01	1.00	1.35
2	11.55	9.59	7.15	2.85	1.70	1.82	2.25	1.93
3	9.95	7.87	8.15	4.05	2.70	2.28	2.70	2.19

Ammons (1952) study, practice seems to affect the stability of a perceptual organization when it is set up by external directions. However, it is possible that practice also affects perceptual organizations when some particular directional set is adopted by *S* himself, either by conscious self-instructions or by sets carried below his level of awareness. In this way, perceptual "habits" may be strengthened below the level of awareness. The ambiguous world, or those aspects of it which are most ambiguous, would thus by practice be perceived in an individual manner.

REFERENCES

ATKINSON, R. C., & AMMONS, R. B. Experiential factors in visual form perception: II. Latency as a function of repetition. *J. exp. Psychol.*, 1952, **43**, 173–178.

RECHTSCHAFFEN, A., & MEDNICK, S. The autokinetic word technique. *J. abnorm. soc. Psychol.*, 1955, **51**, 346.

(Received August 6, 1960)

Recent studies of interest to JEP readers . . .

PSYCHOLOGICAL MONOGRAPHS

No. 498. **The Information Available in Brief Visual Presentations**

By GEORGE SPERLING. Price $1.00

No. 502. **Sleep Deprivation: Changes in Performance and Physiological Indicants of Activation**

By ROBERT B. MALMO and WALTER W. SURWILLO. Price $1.00

No. 503. **Ability Factors and Component Performance Measures as Predictors of Complex Tracking Behavior**

By JAMES F. PARKER, JR. and EDWIN A. FLEISHMAN. Price $1.50

No. 504. **Some Variables Affecting Pattern Conception**

By EVA I. SHIPSTONE. Price $1.50

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street, N.W.
Washington 6, D. C.